

# KLF4 is involved in the organization and regulation of pluripotency-associated three-dimensional enhancer networks

Dafne Campigli Di Giammartino<sup>1,9</sup>, Andreas Kloetgen<sup>2,9</sup>, Alexander Polyzos<sup>1,9</sup>, Yiyuan Liu<sup>1,9</sup>, Daleum Kim<sup>1</sup>, Dylan Murphy<sup>1</sup>, Abderhman Abubashem<sup>1,3,4</sup>, Paola Cavaliere<sup>5</sup>, Boaz Aronson<sup>1</sup>, Veevek Shah<sup>1</sup>, Noah Dephoure<sup>5</sup>, Matthias Stadtfeld<sup>1,6</sup>, Aristotelis Tsirigos<sup>1,6,7,8\*</sup> and Effie Apostolou<sup>1\*</sup>

**Cell fate transitions are accompanied by global transcriptional, epigenetic and topological changes driven by transcription factors, as is exemplified by reprogramming somatic cells to pluripotent stem cells through the expression of OCT4, KLF4, SOX2 and cMYC. How transcription factors orchestrate the complex molecular changes around their target gene loci remains incompletely understood. Here, using KLF4 as a paradigm, we provide a transcription-factor-centric view of chromatin reorganization and its association with three-dimensional enhancer rewiring and transcriptional changes during the reprogramming of mouse embryonic fibroblasts to pluripotent stem cells. Inducible depletion of KLF factors in PSCs caused a genome-wide decrease in enhancer connectivity, whereas disruption of individual KLF4 binding sites within pluripotent-stem-cell-specific enhancers was sufficient to impair enhancer-promoter contacts and reduce the expression of associated genes. Our study provides an integrative view of the complex activities of a lineage-specifying transcription factor and offers novel insights into the nature of the molecular events that follow transcription factor binding.**

Cell identity is defined by a unique gene expression program as well as a characteristic epigenetic landscape and three-dimensional (3D) chromatin topology, features that are constantly supervised by a set of transcription factors known as master regulators<sup>1,2</sup>. Although the ability of master regulators to maintain and change cell identity is well established, the underlying molecular mechanisms remain poorly understood.

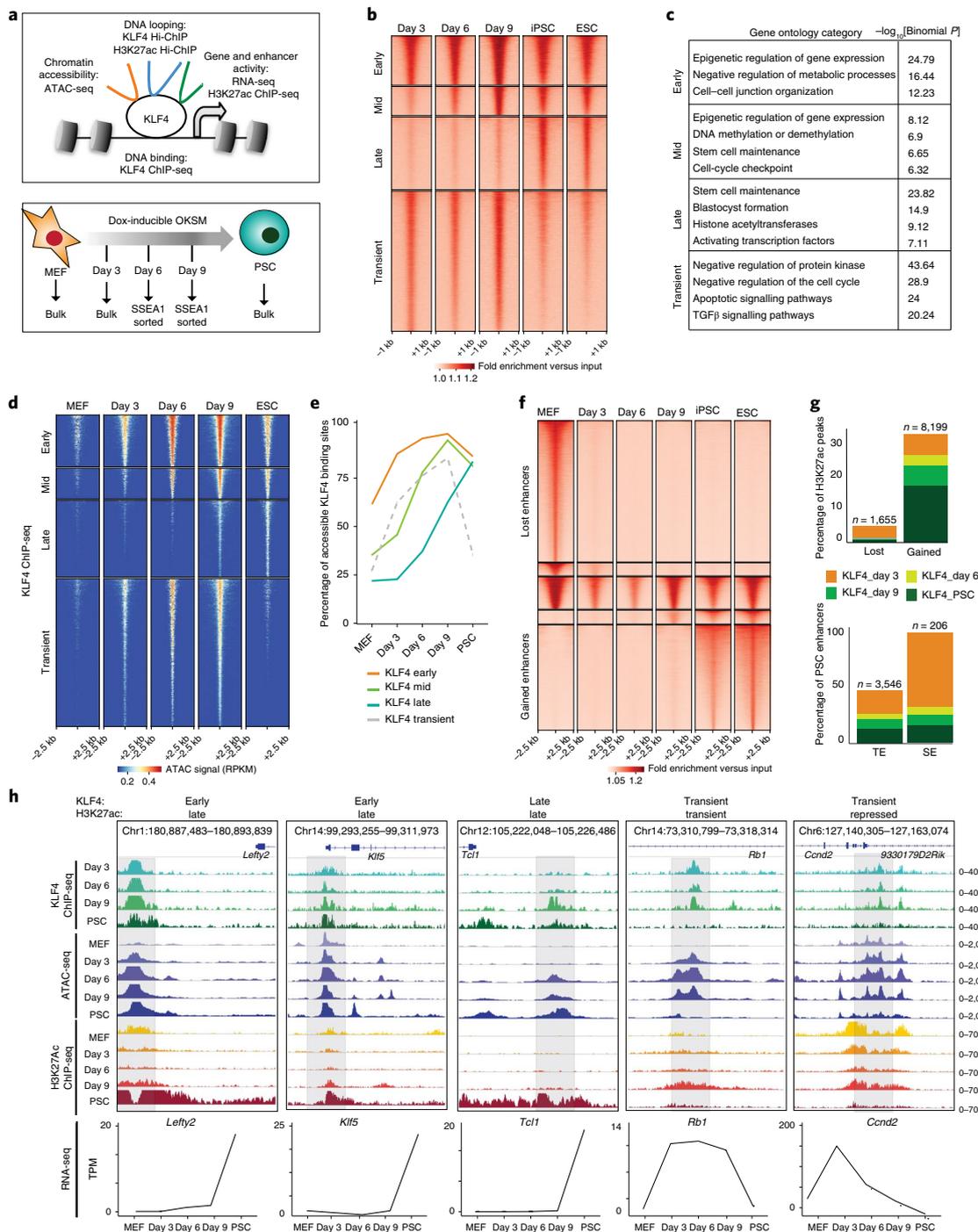
Somatic cell reprogramming into induced pluripotent stem cells (iPSCs) by OCT4, KLF4, SOX2 and cMYC (OKSM) offers a tractable system to study the transcription-factor-driven mechanisms of cell-fate determination<sup>3,4</sup>. The transcriptional and epigenetic changes induced by OKSM expression that result in the erasure of somatic identity and the establishment of pluripotency have been extensively described<sup>5–14</sup>. Recent studies utilizing targeted or global chromatin conformation capture techniques revealed local or large-scale reorganization of the 3D genomic architecture between somatic and pluripotent stem cells (PSCs)<sup>15–21</sup> and a strong association with OKSM binding<sup>15–17,21</sup>, suggesting a potential architectural role of reprogramming transcription factors. The architectural function of KLF4 is further supported by the observations that KLF4 depletion abrogates long-range chromatin contacts at specific genomic loci, such as the *Pou5f1* (*Oct4*) locus in mouse PSCs<sup>18</sup> and the *HOPX* gene in human epidermal keratinocytes<sup>22</sup>. In addition, depletion of the related factor KLF1 disrupts select long-range interactions in

the context of erythropoiesis<sup>23,24</sup>. To test whether OKSM, and in particular KLF4, may orchestrate chromatin architectural changes in a genome-wide manner, we captured the dynamic KLF4-centric topological reorganization and associated molecular alterations during the course of reprogramming mouse embryonic fibroblasts (MEFs) into iPSCs (Fig. 1a, top). Integrative analysis of our results generated a reference map of stage-specific chromatin changes around KLF4-bound loci and established strong links with enhancer rewiring and concordant transcriptional changes. Inducible depletion of KLF factors in PSCs and genetic disruption of KLF4 binding sites within specific PSC enhancers further supported the function of KLF4 as both a transcriptional regulator and a chromatin organizer.

## Results

**KLF4 binding during reprogramming induces chromatin opening and precedes enhancer and gene activation.** We mapped the genome-wide KLF4 binding at different stages of reprogramming using ‘reprogrammable’ MEFs induced with doxycycline (dox)<sup>25</sup> in the presence of ascorbic acid (Fig. 1a, bottom). Under these conditions the resulting iPSCs are molecularly and functionally indistinguishable from embryonic stem cells (ESCs)<sup>26,27</sup> and we used either cell type (referred to as PSCs) as reference points for established pluripotency. Bulk populations were used for the early stages, whereas we sorted SSEA1<sup>+</sup> cells for the mid and late

<sup>1</sup>Sanford I Weill Department of Medicine, Sandra and Edward Meyer Cancer Center, Weill Cornell Medicine, New York, NY, USA. <sup>2</sup>Department of Pathology, NYU School of Medicine, New York, NY, USA. <sup>3</sup>Developmental Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. <sup>4</sup>Weill-Cornell/Rockefeller/Sloan Kettering Tri-Institutional MD-PhD program, New York, NY, USA. <sup>5</sup>Department of Biochemistry, Sandra and Edward Meyer Cancer Center, Weill Cornell Medical College, New York, NY, USA. <sup>6</sup>Skirball Institute of Biomolecular Medicine, Department of Cell Biology and Helen L. and Martin S. Kimmel Center for Biology and Medicine, NYU School of Medicine, New York, NY, USA. <sup>7</sup>Laura and Isaac Perlmutter Cancer Center and Helen L. and Martin S. Kimmel Center for Stem Cell Biology, NYU School of Medicine, New York, NY, USA. <sup>8</sup>Applied Bioinformatics Laboratories, NYU School of Medicine, New York, NY, USA. <sup>9</sup>These authors contributed equally: Dafne Campigli Di Giammartino, Andreas Kloetgen, Alexander Polyzos, Yiyuan Liu. \*e-mail: [aristotelis.tsirigos@nyumc.org](mailto:aristotelis.tsirigos@nyumc.org); [efa2001@med.cornell.edu](mailto:efa2001@med.cornell.edu)



**Fig. 1 | Dynamic KLF4 binding during reprogramming and association with chromatin accessibility and enhancer activity. a**, Schematic illustration of the experimental system and strategy. **b**, Tornado plots of the KLF4 ChIP-seq signals at different reprogramming stages clustered in four different categories: early, mid, late and transient KLF4 binding. The ChIP-seq signals (fold enrichment over input) are shown for 1 kb up- and downstream of the peak centres. **c**, Gene ontology analysis of the early ( $n = 6,275$ ), mid ( $n = 3,712$ ), late ( $n = 9,287$ ) and transient ( $n = 17,891$ ) KLF4 target sites. Significance was calculated using a two-sided binomial test, as provided by the GREAT software. **d**, Tornado plot of the ATAC-seq signal at different reprogramming stages around the KLF4 binding sites. The ATAC-seq signals are shown for 2.5 kb up- and downstream of the peak centres. RPKM, reads per kilobase million. **e**, Line plots showing the percentage of KLF4 early, mid, late and transient targets that overlapped with the ATAC-seq peaks (accessible regions) at each reprogramming stage. **f**, Tornado plot of the H3K27ac ChIP-seq signals showing the MEF, PSC and constant peaks at each reprogramming stage. The ChIP-seq signals (fold enrichment over input) are shown for 2.5 kb up- and downstream of the peak centre. **g**, Percentage of lost (MEF) or gained (PSC) H3K27ac peaks (top) and PSC enhancers (TE) or superenhancers<sup>34</sup> (SE; bottom) that overlap with the KLF4 binding sites at any reprogramming stage. The exact number of overlapping H3K27ac peaks or PSC enhancers are shown above each bar. **h**, Examples of genomic regions (see genomic coordinates) that show the different kinetics of KLF4 binding and H3K27ac occupancy during reprogramming. The Integrative Genomics Viewer (IGV) tracks for KLF4 ChIP-seq, ATAC-seq and H3K27ac ChIP-seq at each reprogramming stage are shown and the signal values are indicated on the right. The transcriptional changes of the depicted genes during reprogramming are shown at the bottom, expressed as transcripts per million (TPM). The mean value of two biological replicates is shown. The ChIP-seq and ATAC-seq experiments were performed in  $n = 2$  biological replicates.

stages<sup>14,25,28,29</sup> (Supplementary Fig. 1a). Chromatin immunoprecipitation-sequencing (ChIP-seq) analysis showed two major categories of KLF4 binding sites: sites that were enriched during the intermediate reprogramming stages but weak in PSCs (transient KLF4 targets) and sites that were strongly detected in PSCs (Fig. 1b and Supplementary Table 1). Among the PSC KLF4 binding sites, 30% were already bound on day 3 (early KLF4 targets), whereas the rest were either gradually established during reprogramming (mid KLF4 targets) or enriched only in established PSCs (late KLF4 targets). Genomic annotation based on their chromatin-state classification<sup>8</sup> and gene ontology analysis<sup>30</sup> (Fig. 1c and Supplementary Fig. 1b) showed that the early KLF4 targets are enriched for promoters of genes involved in metabolic processes and cell-junction organization (Fig. 1c and Supplementary Fig. 1b), which is in agreement with previous reports on KLF4 (ref. <sup>14</sup>). The mid and late KLF4 targets included pluripotency-associated enhancers and were enriched for stem cell maintenance genes, such as *Sox2*, *Nanog*, *Esrrb* and *Klf4*. Transient KLF4 targets were enriched for enhancers active in partially reprogrammed cells<sup>8</sup> (Supplementary Fig. 1b) and genes involved in apoptosis, negative regulation of the cell cycle and signalling pathways associated with differentiation (Fig. 1c). Therefore, transient KLF4 binding might be associated with unsuccessful reprogramming and alternative fates induced by OKSM expression<sup>31–33</sup>.

The integration of data from assay for transposase-accessible chromatin using sequencing (ATAC-seq) revealed that approximately 60% of the early KLF4 binding sites were already open in MEFs, suggesting that pre-existing chromatin accessibility partly explains the early binding of KLF4 on these targets (Fig. 1d,e and Supplementary Fig. 1c). In contrast, the majority (>70%) of the mid and late KLF4 targets were characterized by a closed chromatin configuration in MEFs (Fig. 1d,e) and elevated DNA methylation levels<sup>12</sup> (Supplementary Fig. 1d). These domains gained accessibility concomitantly with KLF4 binding at later time points, suggesting the requirement of additional factors for epigenetic remodelling. We also observed a large number of inaccessible regions in MEFs that became occupied by KLF4 on day 3 (~40% of early and ~75% of transient targets; Fig. 1e), indicating that the ability of this transcription factor to access 'closed' sites is context-dependent (Fig. 1d,e). Motif enrichment analysis revealed distinct classes of candidate transcription factors that may synergize with KLF4 to promote its stage-specific binding (Supplementary Fig. 1e).

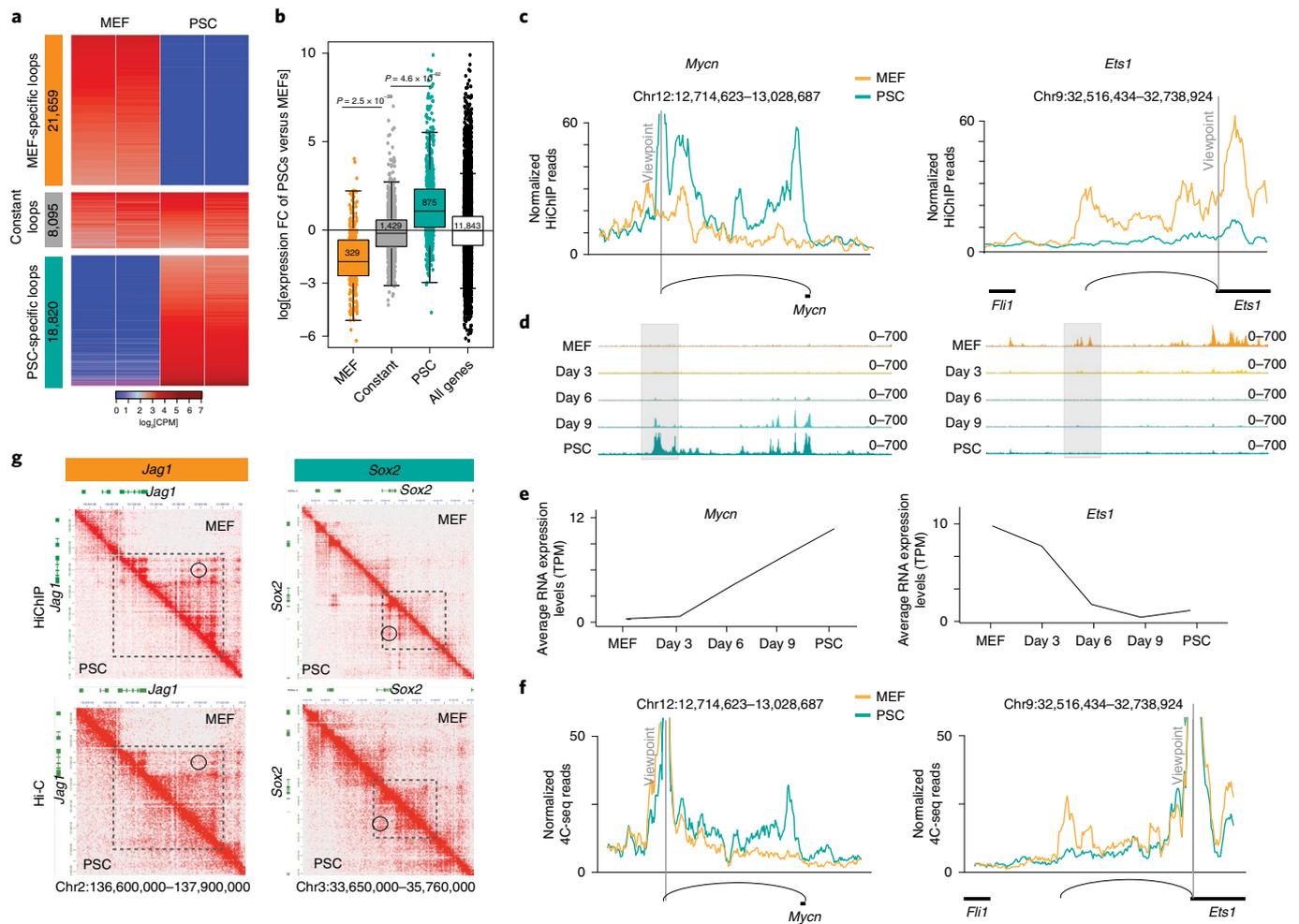
KLF4 has been proposed to function both as an activator and repressor of gene expression<sup>11,14</sup>. ChIP-seq for H3K27 acetylation (H3K27ac) revealed pronounced changes in enhancer activity during iPSC generation associated with KLF4 binding (Fig. 1f and Supplementary Fig. 1f). Although less than 5% of the decommissioned MEF enhancers (regions that lost H3K27ac between MEFs and day 3) were targeted by KLF4, about 35% of the total acquired PSC enhancers and almost the entirety of the so-called superenhancers<sup>34</sup> were bound by KLF4 concomitantly or before H3K27ac (Fig. 1g and Supplementary Table 2). Genes linked to the early, mid, late or transient KLF4 ChIP-seq peaks showed a strong trend for transcriptional upregulation, rather than downregulation, at the respective reprogramming stages (Fig. 1h, Supplementary Fig. 1g,h and Supplementary Table 3), suggesting that KLF4 binding predominantly results in enhancer and gene activation during iPSC generation.

**Long-range enhancer contacts are extensively rewired between MEFs and PSCs in concordance with epigenetic and transcriptional changes.** Targeted (circular chromatin conformation capture (4C) with sequencing, 4C-seq) or global (high-throughput chromosome conformation capture, HiC) chromatin conformation assays have previously demonstrated that the chromatin topology around specific genomic loci and at the scale of compartments

and topologically associating domains (TADs)<sup>35</sup> are drastically reorganized during reprogramming<sup>15,17–21</sup>. As cell-type-specific regulatory loops, such as enhancer–promoter interactions, were underrepresented in these studies, we performed H3K27ac HiChIP<sup>36</sup> to generate high-resolution contact maps around active enhancers in MEFs and PSCs. We called statistically significant interactions using Mango<sup>37</sup> at a resolution of 10 kb, within a maximum range of 2 MB and with at least one anchor overlapping with H3K27ac ChIP-seq peaks (Fig. 2a and Supplementary Fig. 2a; see Methods for details). Differential looping analyses between the normalized read counts (counts per million, CPM) of the union of all called significant loops revealed about 40,000 contacts that were enriched either in MEFs or PSCs (Fig. 2a and Supplementary Table 4). Differential HiChIP contacts were filtered based on *P* values ( $P < 0.1$  and  $\log[\text{fold change (FC)}] > 2$  or  $< -2$ ), which were not corrected for multiple testing. Retrospective validation using edgeR followed by multiple testing correction demonstrated a false-discovery rate (FDR)  $< 0.1$  for the reported differential loops (see Methods). We also identified a group of approximately 9,000 H3K27ac contacts that show constant interaction strength between MEFs and PSCs ( $P > 0.5$ ,  $\log[\text{FC}] < 0.5$  and  $\log[\text{FC}] > -0.5$ ). Mouse-embryonic-fibroblast- and PSC-specific H3K27ac loops correlated with the expression levels of the associated genes (Fig. 2b). These findings demonstrate that H3K27ac HiChIP enables mapping of cell-type-specific enhancer–gene interactions, as exemplified by virtual 4C around the *Mycn* and *Ets1* genes (Fig. 2c). The positions and patterns of the detected chromatin loops at these representative loci are in high concordance with the acquisition or loss of H3K27ac marks and transcriptional changes during reprogramming (Fig. 2d,e). The cell-type-specific nature of these HiChIP-detected interactions was also validated by independent 4C-seq analysis (Fig. 2f).

We performed HiC analysis in MEFs and PSCs to determine whether differential HiChIP contacts reflect chromatin conformation changes or technical bias due to the acquisition or loss of H3K27ac at loop anchors. We observed that only approximately 50% of the HiChIP contacts were also detected in HiC of similar sequencing depth and using the same loop-calling pipeline (Supplementary Fig. 2b). This percentage increased to about 80% when published ultra-resolution HiC data were used<sup>38</sup>, suggesting that sequencing depth is a limiting factor in the detection of HiChIP contacts by HiC. Higher local background in HiC might be another contributing factor, as shown when HiChIP and HiC signals around the *Tbx3* locus were compared (Supplementary Fig. 2c). Contact heat maps further illustrate this point: although both HiChIP and HiC data depict a cell-type-specific configuration around *Jag1* and *Sox2*, several cell-type-specific loops are strongly detected by HiChIP but not by HiC (Fig. 2g). Mouse-embryonic-fibroblast- or PSC-specific HiChIP loops detected by both approaches showed significantly stronger HiC signals in the respective cell type, confirming topological reorganization around these regions (Supplementary Fig. 2d). In agreement with previous reports<sup>36,39</sup>, these observations highlight the increased sensitivity of HiChIP compared with HiC in the detection of cell-type-specific loops.

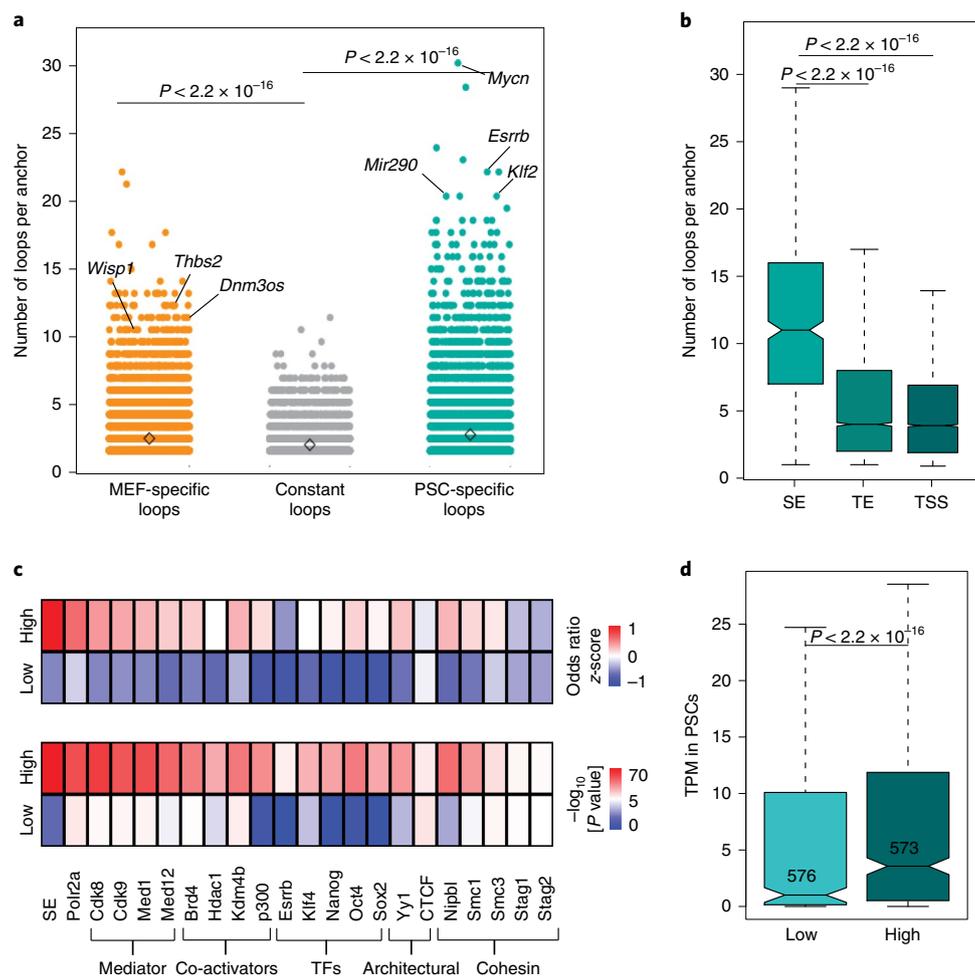
**Complex 3D connectomes in PSCs are associated with strong enhancer activity.** We noticed that specific enhancers and promoters were involved in more than one contact. The degree of connectivity was significantly higher among PSC-specific loops than MEF-specific or constant loops, with hundreds of genomic anchors connecting with ten or more (up to 33) distant genes and/or enhancers (Fig. 3a and Supplementary Fig. 3a). Analysis of HiC data validated the higher connectivity of PSCs compared with MEFs (Supplementary Fig. 3b), possibly reflecting a more plastic chromatin configuration<sup>5,7</sup>.



**Fig. 2 | Characterization of the 3D enhancer connectomes in MEFs and PSCs by H3K27ac HiChIP analysis. a**, Heat map of differential loops detected by H3K27ac HiChIP in MEFs versus PSCs as described in Methods. Numbers indicate the called differential or constant HiChIP loops. **b**, Change in RNA expression between MEFs and PSCs of genes whose promoters were exclusively involved in at least one MEF-specific ( $n = 329$  genes), PSC-specific ( $n = 875$  genes) or constant ( $n = 1,429$  genes) H3K27ac loop. All protein-coding genes ( $n = 11,843$ ) were used as control. Statistics were calculated using an unpaired one-tailed  $t$ -test. The respective minimum and maximum values as well as the 25th, 50th and 75th quartiles are shown for each boxplot. **c–e**, Virtual 4C representation of normalized H3K27ac HiChIP signals around selected viewpoints (*Mycn* enhancer and *Ets1* promoter; **c**). Representative H3K27ac ChIP-seq IGV tracks are shown in **d** for two biological replicates and the average RNA levels expressed from the RNA-seq data of two biological replicates per reprogramming stage are shown in **e**. **f**, Analysis by 4C-seq around the same viewpoints as in **c** validate the presence and cell-type specificity of the HiChIP-detected loops. The 4C-seq signals normalized to the sequencing depth and average across two biological replicates are shown. **g**, HiChIP (top) and HiC (bottom) heat maps generated by Juicebox at a resolution of 10 kb around MEF- (*Jag1*) and PSC-specific (*Sox2*) contacts. Data for both PSCs and MEFs, separated by the diagonal, are shown. The signal indicates the sum of the CPM per matrix (two biological replicates). The dotted squares indicate the regions with a cell-type-specific configuration as detected by both HiC and HiChIP. The circles show examples of cell-type-specific contacts that were detected in the HiChIP and missed in the HiC data.

Among the highest connected regions in PSCs were stem cell regulators, including *Mycn*, *Esrrb* and *mir290*, and PSC super-enhancers<sup>34</sup> (Fig. 3a,b). Highly connected anchors (>4 contacts) preferentially associated with the binding of Pol II, pluripotency transcription factors (including KLF4), mediator complex and transcriptional co-activators (Fig. 3c), and with high transcriptional levels (Fig. 3d). Cohesin subunits and YY1, which can mediate enhancer–promoter loops<sup>40,41</sup>, were also preferentially enriched in highly connected anchors, whereas the classical architectural factor CTCF<sup>42</sup> was not (Fig. 3c). These results suggest that superenhancers and highly expressed genes engage in an increased number of chromatin interactions. Importantly, the number of enhancer contacts showed a poor correlation with the strength of the H3K27ac signal (Supplementary Fig. 3c), suggesting that these observations do not reflect a bias of our HiChIP approach.

**3D-organized enhancer hubs are associated with coordinated cell-type-specific gene expression.** To gain insight into the biological role of the complex enhancer–promoter interactions, we decided to focus on enhancers that establish connections with two or more gene promoters, potentially forming what we refer to as 3D regulatory hubs (or simply enhancer hubs). Genes found within the enhancer hubs were enriched for ‘stem cell maintenance’ categories, including known pluripotency-associated regulators (for example, *Zic2*, *Etv2*, *Lin28a* and *Dnmt3l*; Supplementary Fig. 4a). Hub-connected genes were expressed at significantly higher levels compared with genes with a single-connected enhancer (non-hub genes) and all PSC-expressed genes (Supplementary Fig. 4b). Many of the superenhancers that had been initially assigned to a single gene<sup>34,43</sup> were found to either contact novel distal target genes or form hubs with two or more genes of PSC relevance

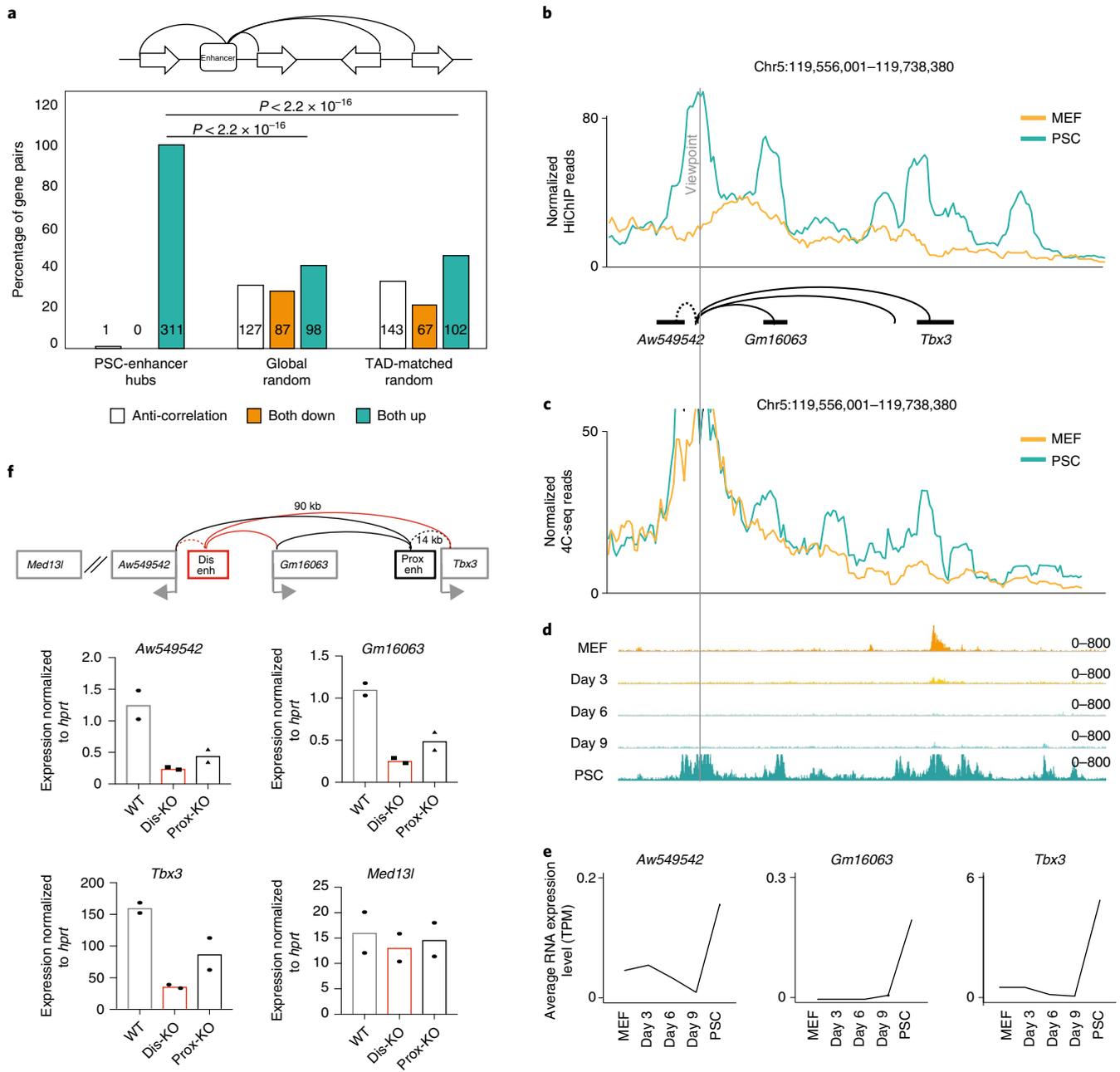


**Fig. 3 | Highly connected enhancers are characterized by specific features. a**, Dot plot showing the number of high-confidence contacts (overall connectivity) around each H3K27ac HiChIP anchor. The number of anchors considered were  $n=26,063$  (MEF-specific), 19,104 (PSC-specific) and 12,316 (constant) loops. **b**, Connectivity of HiChIP anchors containing PSC superenhancers (SE;  $n=474$  anchors), typical enhancers (TE;  $n=5,073$  anchors) and transcription start sites (TSS;  $n=7,493$  anchors) in PSCs. **c**, LOLA enrichment analysis of the enhancer anchors with low ( $n=1,183$  anchors) or high ( $n=1,014$  anchors) connectivity in PSCs using in-house and public ChIP-seq datasets from ESCs (see Methods). The heat maps represent either the  $-\log_{10}[P \text{ value}]$  (bottom), calculated using a two-sided Fisher's exact test, or the z-score of odds ratio (top). TFs, transcription factors. **d**, Expression levels of the genes found in anchors with low ( $n=576$  genes) versus high ( $n=573$  genes) connectivity (expressed in TPM). Significance in **a,b,d** was calculated using a two-sided Wilcoxon rank-sum test. Each boxplot indicates the minimum and maximum values as well as the 25th, 50th and 75th quartiles.

(for example, *Utf1*, *Otx2* and *Nacc1*) and high expression levels (Supplementary Fig. 4a,b). These results expand the previous pool of candidate genes that are regulated by PSC superenhancers<sup>34,44</sup> and suggest that 3D enhancer hubs may coordinate the expression of pluripotency regulators. To test this hypothesis, we calculated the percentage of coregulation of all protein-coding genes that participate in hubs and are differentially expressed between MEFs and PSCs (Supplementary Table 5). This analysis demonstrated a significant overrepresentation of coregulated gene pairs within enhancer hubs compared with control groups of random gene pairs of either similar linear distance (global random) or within the same TADs<sup>35</sup> (TAD-matched random; Fig. 4a) and revealed 311 gene pairs that reside within enhancer hubs and become concordantly upregulated during reprogramming.

To experimentally validate gene coregulation within enhancer hubs, we tested the transcriptional effects of modulating an enhancer hub that contacts two proximal non-coding genes (*Aw549542* and *Gm16063*) and the distal (approximately 90 kb) *Tbx3* gene in a PSC-specific manner (Fig. 4b,c). All enhancers

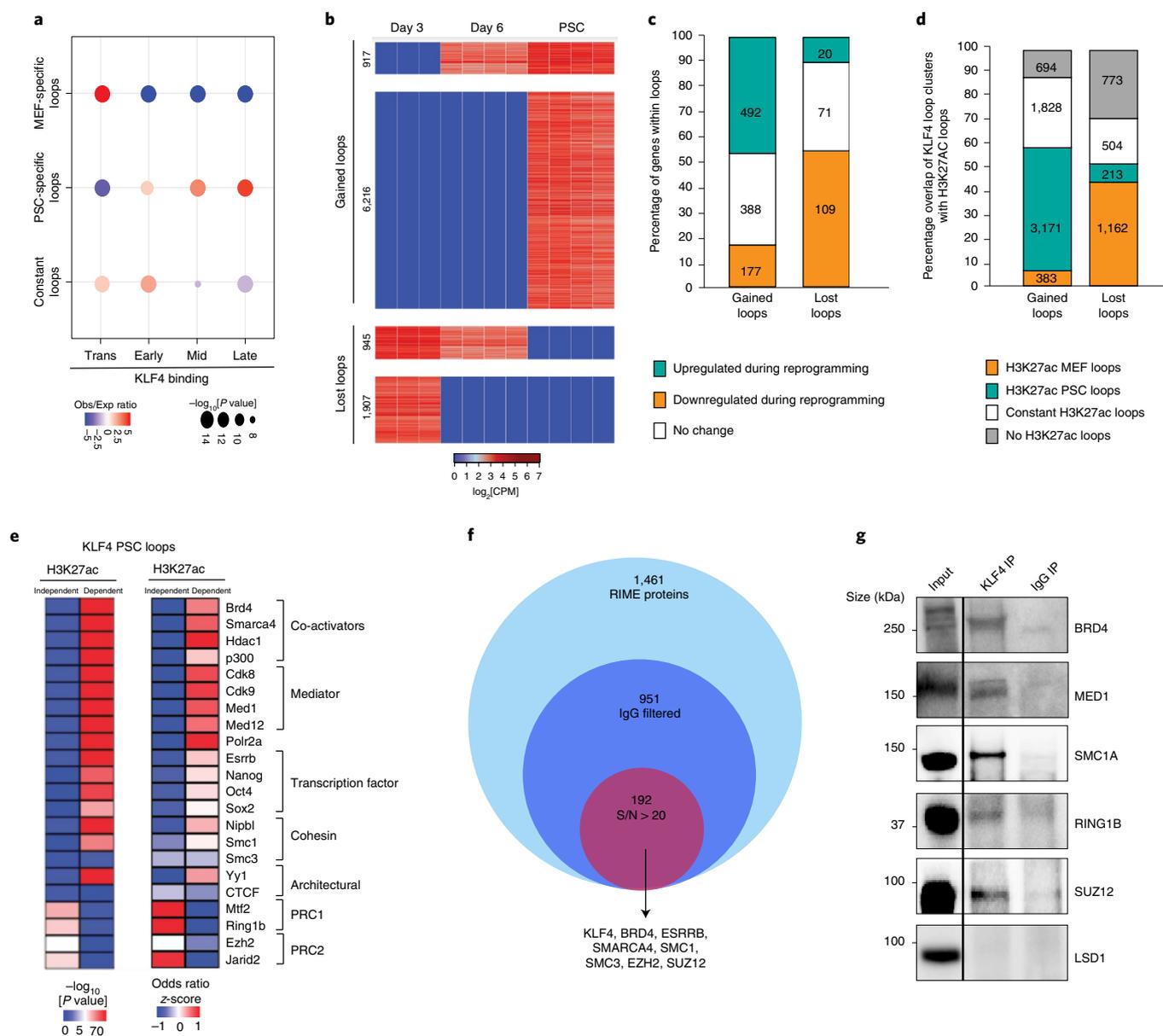
and connected genes within this hub were inactive in MEFs and reprogramming intermediates, and became activated in a coordinated fashion in PSCs (Fig. 4d,e), which was not the case for the *Med13l* gene localized outside the hub (approximately 800 kb; Supplementary Fig. 4c). We deleted the distal *Tbx3* enhancer in PSCs using a deletion of a previously characterized proximal enhancer<sup>45</sup> in the same hub as a reference (Fig. 4f and Supplementary Fig. 4d). The expression of *Tbx3* was severely impaired in homozygous knock-out (KO) clones following the disruption of either enhancer (Dis-KO or Pro-KO), with the distal enhancer showing a stronger effect (Fig. 4f). The RNA levels of the other hub-connected genes (*Gm1603* and *Aw549542*) were also reduced, whereas *Med13l* was unaffected (Fig. 4f). Using dCas9-KRAB<sup>46</sup>, we also targeted a different enhancer that contacts the *Zic2* and *Zic5* genes (Supplementary Fig. 4e), which are co-activated during reprogramming (Supplementary Fig. 4f,g). Silencing of this enhancer (Supplementary Fig. 4h,i) resulted in a significant down-regulation of both genes, whereas the non-hub genes in linear proximity were modestly affected (Supplementary Fig. 4j).



**Fig. 4 | Coregulation of genes within highly interacting enhancer hubs.** **a**, Schematic representation of enhancer hubs interacting with two or more gene promoters (top). Percentage of gene pairs within enhancer hubs that become transcriptionally coregulated (both up- or both downregulated with  $\log_2[FC] \geq 1$  or  $\leq -1$ , adjusted  $P \leq 0.01$ ) or anti-regulated (one up- and one downregulated) between MEFs and PSCs (bottom). Global random or TAD-matched gene pairs were used as controls (see Methods). The individual numbers per group are shown in the respective bars. Statistics were calculated using a two-sided Fisher's exact test. Genes that were not differentially expressed ( $n = 487$ ) were not considered in this analysis. **b**, Example of a newly identified enhancer hub in PSCs. The normalized HiChIP signal around the viewpoint is illustrated as a virtual 4C plot of H3H27ac HiChIP data. **c**, Analysis by 4C-seq around the same viewpoint as in **b**. The average 4C-seq signals normalized to the sequencing depth from two biological replicates are shown. **d**, Representative H3K27ac ChIP-seq IGV tracks from two biological replicates per reprogramming stage. **e**, The RNA-seq signals of the genes within the hub highlight the coordinated upregulation during reprogramming. The average values of two biological replicates are shown. **f**, Experimental strategy (top) for the CRISPR-Cas9-mediated deletion of the *Tbx3* distal (Dis) and proximal (Prox) enhancers (enh) within the hub indicated in panel **b**. Changes in expression (bottom), as determined using RT-qPCR, of *Tbx3*, *Gm16063*, *Aw54954* and a control gene outside the hub (*Med13l*) in CRISPR-Cas9 engineered PSCs carrying homozygous deletions of the distal (Dis-KO;  $n = 2$  independent homozygous KO clones) and proximal (Prox-KO;  $n = 2$  independent homozygous KO clones) *Tbx3* enhancers calculated as a percentage relative to the WT ( $n = 2$  independent clones after empty-vector targeting).  $P$  values were calculated using an unpaired one-tailed  $t$ -test (see also Supplementary Fig. 8 and Supplementary Table 11).

**KLF4-centred chromatin reorganization during reprogramming associates with enhancer rewiring and transcriptional changes of target genes.** The integration of H3K27ac HiChIP and KLF4

ChIP-seq data demonstrated that early, mid and late KLF4 targets (see Fig. 1b) enriched for PSC-specific H3K27ac interactions, whereas MEF-specific contacts enriched for transient KLF4 binding



**Fig. 5 | Chromatin reorganization around KLF4 binding sites during reprogramming associates with enhancer rewiring and requires additional cofactors.**

**a**, Relative enrichment of MEF-specific ( $n = 31,087$ ), PSC-specific ( $n = 24,203$ ) and constant ( $n = 9,389$ ) H3K27ac HiChIP loops that overlapped with KLF4 ChIP-seq peaks, which were clustered as early ( $n = 6,275$ ), mid ( $n = 3,712$ ), late ( $n = 9,287$ ) and transient ( $n = 17,891$ ; see Fig. 1b). The size of the dots indicate the  $P$  values (calculated using a two-sided Fisher’s exact test) and the colours indicate the ratio of observed (Obs) versus expected (Exp). **b**, Heat map of the differential KLF4 HiChIP contacts during reprogramming (see Methods). **c**, Relative proportion of genes within the gained or lost KLF4 loops that were upregulated, downregulated or unchanged during reprogramming (see Methods). The individual numbers of genes per group are shown in the respective bars. **d**, Percentage of gained or lost KLF4 HiChIP loops that were also detected by H3K27ac HiChIP analysis in MEFs and/or PSCs. The individual numbers of KLF4 HiChIP loops per group are shown in the respective bars. **e**, LOLA enrichment analysis of KLF4 binding sites in PSCs that overlap either with H3K27ac-dependent (detected by both KLF4 and H3K27ac HiChIP;  $n = 11,287$  overlapping KLF4 binding sites) or -independent (detected only by KLF4 HiChIP;  $n = 5,078$  overlapping KLF4 binding sites; see Supplementary Fig. 4c) loops. Select factors that scored as significantly enriched over background are shown. The heat maps represent the  $-\log_{10}[P \text{ value}]$  (left) and the z-score of the odds ratio (right). **f**, Venn diagram showing the number of proteins identified by RIME analysis against KLF4 (three biological replicates) after filtering out proteins detected in the IgG control (two biological replicates). Only proteins detected in KLF4 with a signal-to-noise ratio above 20 were considered. **g**, Immunoprecipitation (IP) using KLF4 antibody or IgG in PSC extracts followed by western blot analysis validated the interaction of KLF4 with select factors (LSD1 was used as a negative control). Images are representative of two independent experiments. The relative position and size of the protein markers is shown (see Supplementary Fig. 8).

(Fig. 5a). These results raise the possibility that KLF4 is involved in 3D enhancer reorganization during reprogramming.

We performed KLF4 HiChIP in early (day 3) and mid (day 6) reprogramming intermediates to directly capture the topologi-

cal changes around the KLF4-occupied sites during iPSC formation. Principle component analysis of all statistically significant interactions distinguished KLF4-bound loops from H3K27ac-marked contacts (Supplementary Fig. 5a). Differential looping

analysis revealed two clusters of KLF4-centred interactions gained in the mid or late reprogramming stages, and two clusters of lost loops detected only in the early or mid stages (Fig. 5b and Supplementary Table 6). These lost KLF4 HiChIP contacts mostly associated with gene repression, whereas gained KLF4 loops correlated with gene activation during reprogramming (Fig. 5c). Accordingly, >40% of the lost KLF4 contacts represented MEF enhancer loops, whereas the majority of the gained KLF4 loops overlapped with PSC enhancer interactions (Fig. 5d). These observations support a role of KLF4 binding in the formation/activation of PSC enhancer loops and abrogation/repression of pre-existing somatic loops.

To understand the relative effect of KLF4 binding and/or looping on gene activation, we clustered enhancer–promoter loops detected by both KLF4 and H3K27ac HiChIP in PSCs as loops that: (1) were bound early by KLF4 and formed early during reprogramming (day 3), (2) bound early but formed late and (3) bound late and formed late (Supplementary Fig. 5b, left). Genes in the first category were upregulated early during reprogramming, whereas genes in the other two categories were activated only at the late reprogramming stages (Supplementary Fig. 5b, right). These results indicate that looping coincides with gene activation, whereas KLF4 binding per se is not always sufficient to establish promoter–enhancer contacts and activate transcription.

**KLF4 binding engages in both activating and repressive loops in PSCs.** About 30% of the dynamic KLF4-centred loops did not associate with any expression changes and did not overlap with enhancer contacts (Fig. 5c,d). Among all of the KLF4-centred loops in PSCs, 74% overlapped with H3K27ac HiChIP contacts (H3K27ac-dependent) and 26% were H3K27ac-independent (Supplementary Fig. 5c). KLF4 binding sites within the H3K27ac-dependent loops were enriched for active enhancer features, such as binding of pluripotency transcription factors, RNA Pol II, cohesin and mediator (Fig. 5e). In contrast, H3K27ac-independent KLF4 anchors were enriched for polycomb repressive complex 1 and 2 (PRC1/2) components, which may mediate looping among repressed or bivalent genes in PSCs<sup>17,47</sup>. Genes within the H3K27ac-independent loops were expressed at significantly lower levels (Supplementary Fig. 5d) and associated with development and lineage specification (Supplementary Fig. 5e). These findings raise the possibility that KLF4 engages in chromatin loops with distinct functions, possibly by interacting with different cofactors.

We performed rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME)<sup>48</sup> in PSCs using either a KLF4 antibody or IgG as control (Fig. 5f) to test the chromatin co-occurrence of KLF4 with computationally predicted cofactors. This method identified 228 high-confidence (FC > 1.5 over IgG and  $P < 0.05$ ) protein partners (Supplementary Table 7). Several predicted cofactors—including components of the cohesin complex, PRC1/2 and co-activators such as BRD4—were detected using RIME. Immunoprecipitation followed by western blot analysis validated the interaction of KLF4 with select candidates (Fig. 5g). These results support the idea that KLF4 participates in both (1) activating chromatin loops enriched in cohesin, co-activators and pluripotency transcription factors, and engaging highly expressed genes involved in cell cycle and stemness (for example, *Nodal*, *Mycn* and *Pou5f1*) and (2) repressive loops mediated by PRC1/2 that engage genes related to differentiation and development (for example, *Hoxd10* and *Bmp4*; Supplementary Fig. 5f).

**Depletion of KLF factors in PSCs disrupts a subset of enhancer loops and the expression of linked genes.** We generated an ESC line for the dox-inducible targeting of *Klf4* by CRISPR–Cas9 to dissect the role of KLF4 in the 3D enhancer connectome of PSCs. The levels of KLF4 protein were successfully reduced 48h after the addition

of dox (Supplementary Fig. 6a), but we noticed an upregulation of *Klf2* and *Klf5*, which encode transcription factors with partially redundant function to KLF4 (ref. 49; Supplementary Fig. 6b). We therefore targeted all three KLF factors using the same conditional system (triple KO, TKO). Shortly after dox induction (24h), when the levels of KLF proteins were successfully reduced but before other pluripotency factors such as NANOG were affected (Supplementary Fig. 6c), we performed H3K27ac HiChIP, ChIP-seq and RNA-seq (Supplementary Table 8). We found 7,024 enhancer contacts that were consistently reduced (lost) in all TKO replicates and 3,488 newly established loops (Fig. 6a). More than 60% of the lost loops were bound by KLF4 (ChIP-seq) on one or both anchors, indicating that their disruption is probably a direct effect of the downregulation of KLF factors (Fig. 6b), and multiconnected hubs and superenhancers were preferentially affected compared with typical enhancers (Supplementary Fig. 6d).

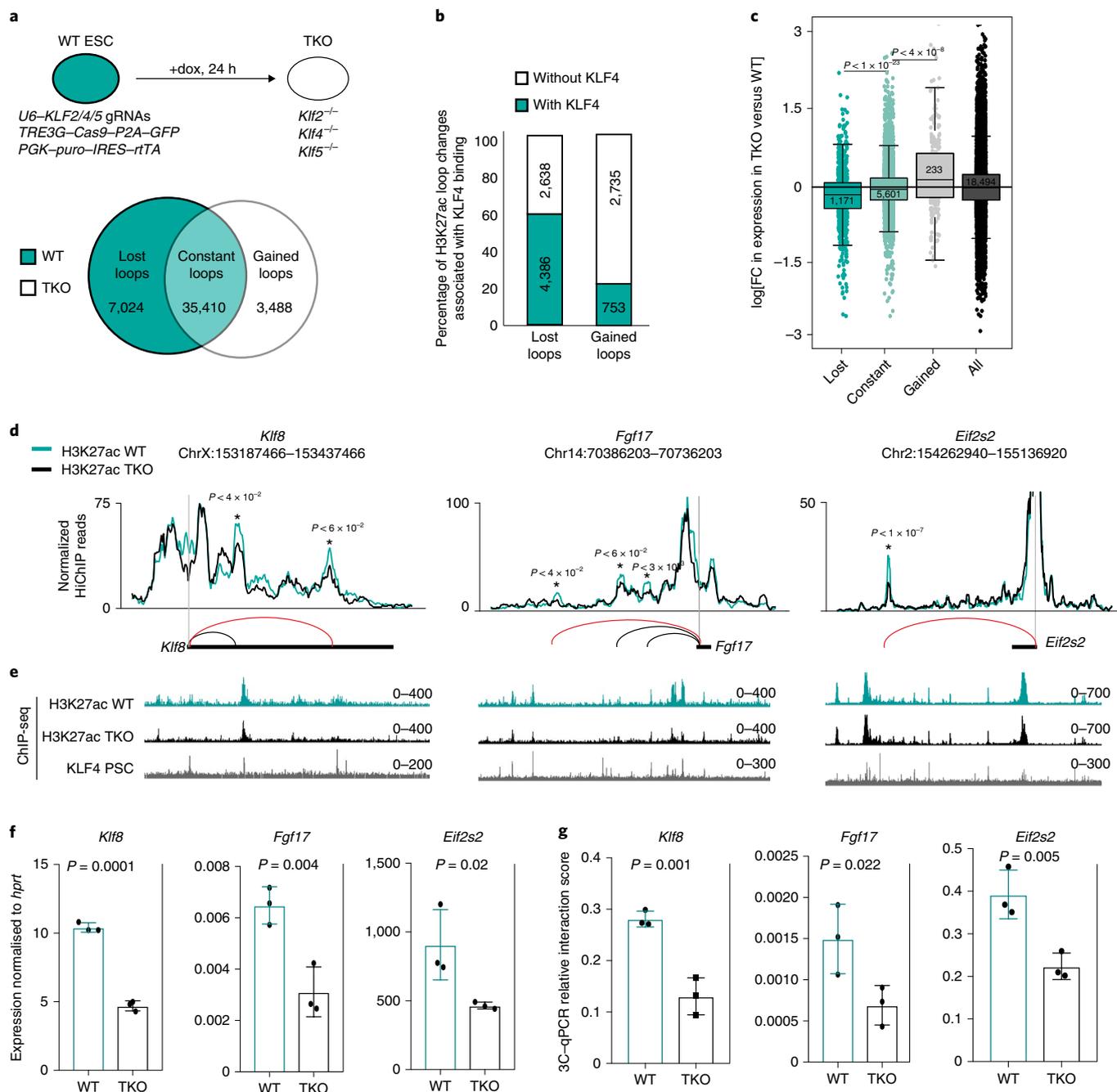
Genes within the lost and gained loops were significantly down- and upregulated, respectively, in TKO compared with wild-type (WT) cells (Fig. 6c). Examples of lost loops involving the *Klf8*, *Fgf17* and *Eif2s2* genes are shown in Fig. 6d,e. The observed reductions in messenger RNA levels and disruption of gene–enhancer contacts in TKO cells were validated by quantitative PCR (qPCR) after reverse transcription (RT–qPCR) and chromosome conformation capture with qPCR (3C–qPCR), respectively (Fig. 6f,g). These results demonstrate that the depletion of KLF factors in PSCs results in the genome-wide abrogation of enhancer contacts and concordant dysregulation of connected genes.

**Disruption of KLF4 binding sites interferes with enhancer looping and transcriptional activation.** We targeted KLF4 binding sites within selected enhancer hubs to ascertain whether KLF4 binding is critical for the architecture and activity of the enhancer–promoter contacts. We initially chose the distal *Tbx3* enhancer, the deletion of which resulted in the downregulation of all three hub-connected genes (Fig. 4f). Contacts of this enhancer with the surrounding genes were detected in PSCs but not in MEFs or reprogramming intermediates (Fig. 7a), which is in concordance with late binding of KLF4 (Fig. 7b) and late transcriptional activation of the entire locus (Fig. 4e). We disrupted the strongest KLF4 binding motif within this enhancer hub (Fig. 7c and Supplementary Fig. 7a–c) and used four homozygous mutant clones with impaired KLF4 binding (Supplementary Fig. 7d) for further characterization. The expression levels of all hub-connected genes (*Aw549542*, *Gm1603* and *Tbx3*) were significantly reduced, whereas a gene outside the hub (*Med13l*) was not affected (Fig. 7d). Consistent with transcriptional downregulation, the long-range contacts between the enhancer hub and its targets were significantly weakened in mutant clones (Fig. 7e), whereas the interaction of *Tbx3* with the proximal enhancer or a KLF4-independent contact in a different genomic region remained unaffected (Fig. 7e).

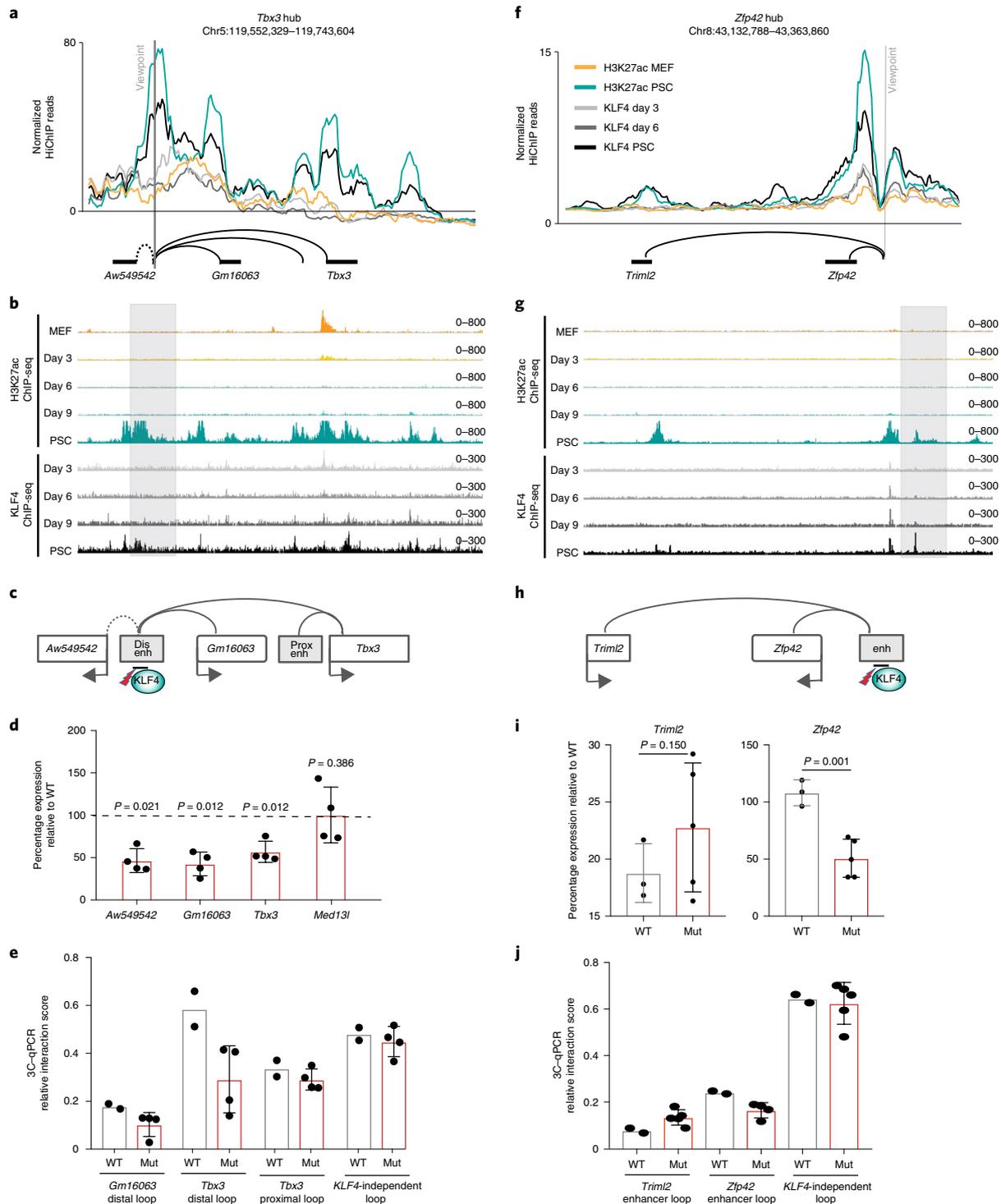
We also mutated a KLF4 binding site within the previously described *Zfp42* superenhancer<sup>50</sup>, which contacts both *Zfp42* and the distal (approximately 150 kb) *Triml2* gene in PSCs (Fig. 7f,g). Homozygous mutant ESCs showed downregulation of *Zfp42* expression and a concordant reduction in the frequency of enhancer–*Zfp42* promoter contact (Fig. 7h–j). Intriguingly, the expression levels of *Triml2* remained unaffected in the mutant clones and the connection with the enhancer seemed to be even stronger (Fig. 7i,j), suggesting that KLF-dependent and -independent mechanisms may regulate the looping and activity of the same enhancer on different genes. Together, these results provide evidence for a dual role of KLF4 as a transcriptional regulator and chromatin organizer in PSCs.

## Discussion

Here, we describe the genome-wide dynamics of KLF4 binding and its effects on chromatin accessibility, enhancer activity, gene



**Fig. 6 | Inducible depletion of KLF proteins induces 3D enhancer reorganization and concordant transcriptional changes.** **a**, Experimental strategy used to knock out KLF2, KLF4 and KLF5 proteins in ESCs using a dox-inducible CRISPR–Cas9 construct (top; see also Supplementary Fig. 8). Venn diagram showing the number of H3K27ac HiChIP loops that were gained or lost ( $P < 0.05$  and  $FC > 1.5$  or  $< -1.5$ ) or remained constant ( $\log[FC] > -0.5$  and  $< 0.5$  and  $P > 0.5$ ) in TKO ESCs compared with uninduced ESCs (bottom; see Methods). **b**, Percentage of gained or lost H3K27ac HiChIP contacts, whose anchors do or do not overlap with KLF4 ChIP-seq peaks in TKO versus WT PSCs. The individual numbers of HiChIP contacts per group are shown in the respective boxes. **c**, Changes in the RNA expression of genes within the anchors of the H3K27ac HiChIP contacts. All protein-coding genes were used as a control. The respective gene numbers are shown in the boxes. Statistics were calculated using an unpaired one-tailed *t*-test. Each boxplot indicates the minimum and maximum values as well as the 25th, 50th and 75th quartiles. **d**, Examples of H3K27ac lost interactions in TKO versus WT ESC, as identified by H3K27ac HiChIP. Normalized H3K27ac HiChIP signals are illustrated in a virtual 4C format around the viewpoints. Statistics were calculated using the R package edgeR (see Methods) using  $n = 2$  biological replicates. **e**, Representative H3K27ac and KLF4 ChIP-seq tracks from two biological replicates around each of the genomic regions indicated in **d**. **f**, Mean changes in the expression of *Klf8*, *Fgf17* and *Eif2s2*, as determined by RT-qPCR, in WT ( $n = 3$  biological replicates) and TKO ( $n = 3$  biological replicates) PSCs relative to the *hprt* levels. The *P* values were calculated using an unpaired one-tailed *t*-test. **g**, Analysis by 3C-qPCR validating the reduced contact frequency between the *Klf8*, *Fgf17* and *Eif2s2* promoters and their respective enhancers (marked with a red line in panel **d**) in TKO compared with WT ESCs ( $n = 3$  biological replicates). An unpaired one-tail *t*-test was used to determine the *P* values. The error bars indicate the s.d. (Source data in Supplementary Table 11).



**Fig. 7 | Disruption of the KLF4 binding site within *Tbx3* and *Zfp42* enhancers induces looping abrogation and downregulation of linked genes in PSCs.**

**a, b.** Normalized KLF4 and H3K27ac HiChIP signals illustrated as virtual 4C line plots around the *Tbx3* distal enhancer hub (**a**; see also Fig. 4b–f). The signals show the average CPM per condition for two (H3K27ac HiChIP), three (Klf4 day 3 HiChIP) and four (Klf4 day 6/PSC HiChIP) biological replicates. Representative ChIP-seq tracks from two biological replicates are shown in **b**. **c.** CRISPR–Cas9 targeting strategy to disrupt KLF4 binding motifs (Mut) within the distal *Tbx3* enhancer. **d.** Changes in expression of the hub-associated genes, as determined by RT–qPCR. *Med13l* was used as a control gene outside the hub. The bar plots show the mean value calculated as a percentage relative to the WT after normalization relative to the *hprt* mRNA levels from two independent experiments. A paired one-tailed *t*-test was used to determine significance relative to the WT ( $n = 4$  PSC clones). The error bars indicate the s.d. among  $n = 4$  PSC clones carrying homozygous mutations of the KLF4 binding motif (Mut). **e.** Analysis by 3C–qPCR showing the relative interaction frequency of the *Tbx3* distal enhancer with the promoters of linked genes in WT ( $n = 2$  PSC clones) and mutant ( $n = 4$  PSC clones). The bar plots show the mean values and the error bars indicate the s.d. **f–j.** Representation, analysis and functional validation of the *Zfp42* enhancer hub similarly to **a–e** for the *Tbx3* hub. The same normalizations and statistical tests (unpaired one-tailed *t*-test) were applied. **i.**  $n = 3$  WT PSC clones and  $n = 5$  KLF4 mutant clones. **j.**  $n = 2$  WT and  $n = 5$  mutant clones. See Supplementary Fig. 8 and Supplementary Table 11.

expression and 3D enhancer organization during iPSC reprogramming and in established PSCs. Our data suggest that the relationship between KLF4 binding and gene and enhancer activation at some loci is dependent on pre-existing chromatin accessibility or DNA methylation and/or the availability of additional transcription factors and cofactors, such as ESRRB or NANOG<sup>8,12,13</sup>. Nevertheless, KLF4 also binds to chromatin regions that are inaccessible and highly methylated in somatic cells, which is in agreement with its documented ability to act as a pioneer factor<sup>9,51,52</sup> and/or its cooperative binding with other reprogramming transcription factors<sup>8</sup>.

Studies utilizing 4C or HiC have offered important insights into topological reorganization during somatic cell reprogramming but do not capture the dynamic assembly and disassembly of cell-type-specific enhancer contacts<sup>15,18,21</sup>. HiChIP allowed us to generate reference maps of regulatory loops in MEFs and PSCs, and revealed an extensive rewiring of enhancer connectomes during reprogramming. The 4C-seq and HiC experiments validated the cell-type-specific nature of the HiChIP interactions but also revealed technical biases and limitations for each approach, thus highlighting the need for a systematic comparison of different conformational assays and analytical tools.

We have uncovered a set of highly connected enhancers, which communicate with strongly expressed cell-type-specific genes, supporting the idea that high interactivity might be an inherent characteristic of the regulatory elements critical for cell identity<sup>53,54</sup>. In addition, we identified cell-type-specific enhancers, including many superenhancers, which frequently interact with two or more coregulated genes, which provides support for a role for such hubs in coordinating target gene activation, as previously shown in different contexts<sup>55</sup>. Accordingly, the deletion or inactivation of enhancer hubs resulted in coordinated downregulation of all connected genes without affecting the neighbouring non-hub genes. Recently developed technologies that capture multiway interactions<sup>56–59</sup> will enable the investigation of the extent to which the enhancer hubs represent multiple contacts occurring in the same cell and on the same allele.

Increasing evidence suggests that transcription factors mediate chromatin contacts in different cellular contexts (reviewed in ref. <sup>60</sup>), although the underlying mechanisms and the relationship between transcription factor binding and the topological and transcriptional changes remain elusive. Encouraged by previous studies reporting potential architectural functions for KLF4-protein members<sup>18,23,24</sup>, we captured the dynamic chromatin reorganization around KLF4 binding sites during iPSC formation in a direct and genome-wide manner. This revealed that KLF4 binding is associated with the de novo establishment of long-range enhancer contacts and promotes the transcriptional upregulation of linked genes. KLF4 binding is not always sufficient for loop formation and gene activation, which suggests that additional architectural factors and coregulators are required. Our computational and proteomics analyses revealed distinct sets of candidate cofactors, which may interact with KLF4 in the context of activating enhancer loops or repressive/poised loops, respectively. How do these proteins work together to form 3D chromatin contacts? The recruitment of architectural cofactors capable of tethering distal DNA elements is a plausible scenario, supported by the interaction between KLF4 and cohesin subunits<sup>43</sup>. Another possibility is that the formation of topological assemblies, such as 3D enhancer hubs or polycomb bodies<sup>17,47,53,58,61,62</sup>, is the result of 'self-organization' due to a homotypic chromatin state or multiprotein condensation. KLF4 and validated cofactors, such as mediator and BRD4, are characterized by extensive intrinsically disordered regions, which have been proposed to promote multivalent interactions and the formation of subnuclear condensates<sup>63,64</sup>.

Our study describes the functional role of KLF4 in the genome-wide organization and regulation of enhancer contacts, extending previous descriptions of the involvement of this factor in the

maintenance of selected chromatin loops<sup>18,22</sup>. In addition to the global topological effects triggered by KLF4 protein depletion, we show that targeting KLF4 binding sites within specific enhancer hubs can be sufficient to disrupt enhancer–promoter contacts and induce transcriptional downregulation. Systematic functional interrogation of KLF4-bound enhancer hubs should enable a deeper understanding of the molecular mechanisms of topological organization and the establishment of new criteria for the identification and functional prioritization of regulatory nodes critical for PSC identity.

### Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at <https://doi.org/10.1038/s41556-019-0390-6>.

Received: 20 July 2018; Accepted: 16 August 2019;

Published online: 23 September 2019

### References

- Young, R. A. Control of the embryonic stem cell state. *Cell* **144**, 940–954 (2011).
- Natoli, G. Maintaining cell identity through global control of genomic organization. *Immunity* **33**, 12–24 (2010).
- Graf, T. Historical origins of transdifferentiation and reprogramming. *Cell Stem Cell* **9**, 504–516 (2011).
- Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
- Apostolou, E. & Hochedlinger, K. Chromatin dynamics during cellular reprogramming. *Nature* **502**, 462–471 (2013).
- Apostolou, E. & Stadtfeld, M. Cellular trajectories and molecular mechanisms of iPSC reprogramming. *Curr. Opin. Genet. Dev.* **52**, 77–85 (2018).
- Di Giandomartino, D. C. & Apostolou, E. The chromatin signature of pluripotency: establishment and maintenance. *Curr. Stem Cell. Rep.* **2**, 255–262 (2016).
- Chronis, C. et al. Cooperative binding of transcription factors orchestrates reprogramming. *Cell* **168**, 442–459 e420 (2017).
- Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151**, 994–1004 (2012).
- Chen, J. et al. Hierarchical Oct4 binding in concert with primed epigenetic rearrangements during somatic cell reprogramming. *Cell Rep.* **14**, 1540–1554 (2016).
- Sridharan, R. et al. Role of the murine reprogramming factors in the induction of pluripotency. *Cell* **136**, 364–377 (2009).
- Knaupp, A. S. et al. Transient and permanent reconfiguration of chromatin and transcription factor occupancy drive reprogramming. *Cell Stem Cell* **21**, 834–845 e836 (2017).
- Li, D. et al. Chromatin accessibility dynamics during iPSC reprogramming. *Cell Stem Cell* **21**, 819–833 e816 (2017).
- Polo, J. M. et al. A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* **151**, 1617–1632 (2012).
- Apostolou, E. et al. Genome-wide chromatin interactions of the *Nanog* locus in pluripotency, differentiation, and reprogramming. *Cell Stem Cell* **12**, 699–712 (2013).
- de Wit, E. et al. The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature* **501**, 227–231 (2013).
- Denholtz, M. et al. Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. *Cell Stem Cell* **13**, 602–616 (2013).
- Wei, Z. et al. Klf4 organizes long-range chromosomal interactions with the Oct4 locus in reprogramming and pluripotency. *Cell Stem Cell* **13**, 36–47 (2013).
- Krijger, P. H. et al. Cell-of-origin-specific 3D genome structure acquired during somatic cell reprogramming. *Cell Stem Cell* **18**, 597–610 (2016).
- Beagan, J. A. et al. Local genome topology can exhibit an incompletely rewired 3D-folding state during somatic cell reprogramming. *Cell Stem Cell* **18**, 611–624 (2016).
- Stadhouders, R. et al. Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. *Nat. Genet.* **50**, 238–249 (2018).
- Rubin, A. J. et al. Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. *Nat. Genet.* **49**, 1522–1528 (2017).

23. Drissen, R. et al. The active spatial organization of the  $\beta$ -globin locus requires the transcription factor EKLf. *Genes Dev.* **18**, 2485–2490 (2004).
24. Schoenfelder, S. et al. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat. Genet.* **42**, 53–61 (2010).
25. Stadtfeld, M., Maherali, N., Borkent, M. & Hochedlinger, K. A reprogrammable mouse strain from gene-targeted embryonic stem cells. *Nat. Methods* **7**, 53–55 (2010).
26. Stadtfeld, M. et al. Ascorbic acid prevents loss of *Dlk1-Dio3* imprinting and facilitates generation of all-iPS cell mice from terminally differentiated B cells. *Nat. Genet.* **44**, 398–405 (2012).
27. Stadtfeld, M. et al. Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. *Nature* **465**, 175–181 (2010).
28. Zunder, E. R., Lujan, E., Goltsev, Y., Wernig, M. & Nolan, G. P. A continuous molecular roadmap to iPSC reprogramming through progression analysis of single-cell mass cytometry. *Cell Stem Cell* **16**, 323–337 (2015).
29. Stadtfeld, M., Maherali, N., Breault, D. T. & Hochedlinger, K. Defining molecular cornerstones during fibroblast to iPSC cell reprogramming in mouse. *Cell Stem Cell* **2**, 230–240 (2008).
30. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
31. Guo, L. et al. Resolving cell fate decisions during somatic cell reprogramming by single-cell RNA-Seq. *Mol. Cell* **73**, 815–829 (2019).
32. Allison, T. F. et al. Identification and single-cell functional characterization of an endodermally biased pluripotent substate in human embryonic stem cells. *Stem Cell Rep.* **10**, 1895–1907 (2018).
33. Schiebinger, G. et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell* **176**, 928–943 (2019).
34. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
35. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
36. Mumbach, M. R. et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* **13**, 919–922 (2016).
37. Phanstiel, D. H., Boyle, A. P., Heidari, N. & Snyder, M. P. Mango: a bias-correcting ChIA-PET analysis pipeline. *Bioinformatics* **31**, 3092–3098 (2015).
38. Bonev, B. et al. Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572 (2017).
39. Mumbach, M. R. et al. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.* **49**, 1602–1612 (2017).
40. Weintraub, A. S. et al. YY1 is a structural regulator of enhancer-promoter loops. *Cell* **171**, 1573–1588 (2017).
41. Beagan, J. A. et al. YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Res.* **27**, 1139–1152 (2017).
42. Gomez-Diaz, E. & Corces, V. G. Architectural proteins: regulators of 3D genome organization in cell fate. *Trends Cell Biol.* **24**, 703–711 (2014).
43. Downen, J. M. et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374–387 (2014).
44. Novo, C. L. et al. Long-range enhancer interactions are prevalent in mouse embryonic stem cells and are reorganized upon pluripotent state transition. *Cell Rep.* **22**, 2615–2627 (2018).
45. Kearns, N. A. et al. Functional annotation of native enhancers with a Cas9–histone demethylase fusion. *Nat. Methods* **12**, 401–403 (2015).
46. Larson, M. H. et al. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* **8**, 2180–2196 (2013).
47. Mas, G. & Di Croce, L. The role of Polycomb in stem cell genome architecture. *Curr. Opin. Cell Biol.* **43**, 87–95 (2016).
48. Mohammed, H. et al. Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. *Nat. Protoc.* **11**, 316–326 (2016).
49. Jiang, J. et al. A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nat. Cell Biol.* **10**, 353–360 (2008).
50. Zhang, S. et al. Epigenetic regulation of REX1 expression and chromatin binding specificity by HMGNs. *Nucleic Acids Res.* **47**, 4449–4461 (2019).
51. Soufi, A. et al. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555–568 (2015).
52. Sardina, J. L. et al. Transcription factors drive Tet2-mediated enhancer demethylation to reprogram cell fate. *Cell Stem Cell* **23**, 727–741 (2018).
53. Schmitt, A. D. et al. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* **17**, 2042–2059 (2016).
54. Huang, J. et al. Dissecting super-enhancer hierarchy based on chromatin interactions. *Nat. Commun.* **9**, 943 (2018).
55. Fukaya, T., Lim, B. & Levine, M. Enhancer control of transcriptional bursting. *Cell* **166**, 358–368 (2016).
56. Oudelaar, A. M. et al. Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains. *Nat. Genet.* **50**, 1744–1751 (2018).
57. Zheng, M. et al. Multiplex chromatin interactions with single-molecule precision. *Nature* **566**, 558–562 (2019).
58. Allahyar, A. et al. Enhancer hubs and loop collisions identified from single-allele topologies. *Nat. Genet.* **50**, 1151–1160 (2018).
59. Olivares-Chauvet, P. et al. Capturing pairwise and multi-way chromosomal conformations using chromosomal walks. *Nature* **540**, 296–300 (2016).
60. Stadhouders, R., Filion, G. J. & Graf, T. Transcription factors and 3D genome conformation in cell-fate decisions. *Nature* **569**, 345–354 (2019).
61. Schoenfelder, S. et al. Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat. Genet.* **47**, 1179–1186 (2015).
62. Jiang, T. et al. Identification of multi-loci hubs from 4C-seq demonstrates the functional importance of simultaneous interactions. *Nucleic Acids Res.* **44**, 8714–8725 (2016).
63. Boija, A. et al. Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* **175**, 1842–1855 (2018).
64. Cho, W. K. et al. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018).

## Acknowledgements

We are grateful to A. Melnick and the members of the Apostolou, Tsirigos and Stadtfeld laboratories for critical reading of the manuscript. W.W. applied an unpaired, one-sided test also want to thank Z. Chen and the Biostatistics and Epidemiology Consulting Service for the advice and final evaluation on the statistical tools and analyses and L. Dow for sharing the CRISPR–Cas9 vectors. D.C.G. was supported by the New York Stem Cell Foundation and the Family-Friendly Postdoctoral Initiative at Weill Cornell Medicine. A.A. is supported by a Medical Scientist Training Program grant from the National Institute of General Medical Sciences of the National Institutes of Health (NIH) under award number T32GM007739 to the Weill Cornell/Rockefeller/Sloan Kettering Tri-Institutional MD-PhD Program. A.T. is supported by the American Cancer Society (grant no. RSG-15-189-01-RMC), the Leukemia and Lymphoma Society and the St. Baldrick's Foundation. E.A. is supported by the NIH Director's New Innovator Award (grant no. DP2DA043813) and the Tri-Institutional Stem Cell Initiative by the Starr Foundation.

## Author contributions

E.A. conceived, designed and supervised the study, and wrote the manuscript together with D.C.D.G., with help from all of the authors. D.C.D.G. performed all of the experiments with help from D.K. and V.S. A.K. and A.P. performed all HiChIP, HiC and integrative computational analyses under the guidance of A.T. Y.L. performed the initial ChIP-seq, RNA-seq and ATAC-seq analyses. D.M. performed the HiC and CRISPRi experiments using a stable dCas9–KRAB ESC line generated by B.A. A.A. performed the RIME experiments and iPSC ChIP-seq. P.C. and N.D. ran and analysed the RIME results. M.S. provided the reprogrammable cells and guidance on the reprogramming experiments.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41556-019-0390-6>.

**Correspondence and requests for materials** should be addressed to A.T. or E.A.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

## Methods

**Cell lines, culture conditions and reprogramming experiments.** Mouse V6.5 ESCs were cultured on irradiated feeder cells in KO-DMEM media (Invitrogen) supplemented with 15% heat-inactivated fetal bovine serum, GlutaMAX, penicillin-streptomycin, non-essential amino acids,  $\beta$ -mercaptoethanol and  $1,000 \text{ U ml}^{-1}$  leukaemia inhibitory factor, with or without the presence of 2i ( $1 \mu\text{M}$  MEK inhibitor (Stemgent, 04-0006) and  $3 \mu\text{M}$  GSK3 inhibitor (Stemgent, 04-0004)).

Mouse embryonic fibroblasts were isolated from a 'reprogrammable' mouse harbouring a polycistronic OKSM cassette in the *Colla1* locus and M2rtTA in the *Rosa26* locus<sup>25</sup>. The cells were reprogrammed in the presence of  $1 \mu\text{g ml}^{-1}$  dox and  $50 \mu\text{g ml}^{-1}$  ascorbic acid, and cultured in ESC medium as described above. The cells were collected at the indicated time points.

**Animal experiments.** Derivation, handling and genotyping of reprogrammable mice (JAX011001) were described previously<sup>25</sup>. Mouse embryonic fibroblasts were isolated from male and female embryos at day E13.5. All of the animal experiments were in accordance with the guidelines of the Institutional Animal Care and Use Committee at Weill Cornell Medicine under the protocol 2014-0044.

**Lentiviral production and infection.** 293T cells were transfected with overexpression constructs along with the packaging vectors VSV-G and Delta8.9 using PEI reagent (PEI MAX, Polyscience, 24765-2). The supernatant was collected after 48 and 72 h, and the virus was concentrated using polyethylglycol (Sigma, P4338). V6.5 cells were infected in medium containing  $5 \mu\text{g ml}^{-1}$  polybrene (Millipore, TR-1003-G), followed by centrifugation at 2,100 r.p.m. for 90 min at 32 °C.

**FACS.** We used magnetic microbeads conjugated to anti-SSEA1 antibody (see Supplementary Table 10) for the isolation of SSEA1<sup>+</sup> cells from reprogramming intermediates at days 6 and 9 as per the manufacturer's instructions. The SSEA positive and negative fractions were then stained for FACS analysis with an anti-Thy1 antibody conjugated to pacific blue fluorophore and anti-SSEA1 antibody conjugated to APC fluorophore.

**Generation, selection and validation of KO cell lines.** Genomic RNA (gRNA) was cloned into the px458 vector (Addgene, cat. no. 48138) using the BbsI restriction enzyme. V6.5 ESCs ( $0.3 \times 10^6$ ) were transfected using 2  $\mu\text{g}$  Left-*Tbx3*-plasmid and 2  $\mu\text{g}$  Right-*Tbx3*-plasmid (for the *Tbx3* enhancer deletions), 4  $\mu\text{g}$  *Tbx3*-KLF4mut-vector (mutation of the KLF4 binding site within the *Tbx3* distal enhancer) or 4  $\mu\text{g}$  KLF4-Zfp42mut (mutation of the KLF4 binding site within the *Zfp42* enhancer). The DNA was pre-mixed with 50  $\mu\text{l}$  media with no additives, and 10  $\mu\text{l}$  Lipofectamine 2000 (Invitrogen, cat. no. 11668019) was pre-mixed with 50  $\mu\text{l}$  media with no additives. The two tubes were combined after 5 min and incubated at room temperature for another 20 min. The cells were then added to the solution and plated on a gelatinized 12-well plate. Single GFP-positive cells were sorted by FACS into 96-well plates 48 h post transfection. Genotyping of the single-cell colonies was performed using a three-primer strategy (for deletions) or by surveyor with T7 (for insertion or deletion mutations). Four (*Tbx3* hub) or five (*Zfp42* hub) colonies with homozygous mutations (or WT colonies as controls) were expanded and used for RT-qPCR and 3C experiments. All of the gRNA, 3C and RT-qPCR primers are described in Supplementary Table 9.

**CRISPRi of *Zic2/5* enhancer.** V6.5 cells were infected with lentiviruses harbouring the pHR-SFFV-dCas9-BFP-KRAB vector (Addgene, cat. no. 46911), in which the SFFV promoter was replaced with an Efla promoter. Cells expressing BFP were selected by three rounds of FACS sorting. The resulting V6.5 cells, stably expressing the dCas9-KRAB, were then infected with a lentivirus harbouring the pLKO5.GRNA.EFS.PAC vector (Addgene, cat. no. 57825) containing two gRNAs targeting the *Zic2/5* enhancer. The cells were selected with puromycin (LifeTech, K210015) for two days and subsequently collected for RT-qPCR analysis. The gRNA and RT-qPCR primers are described in Supplementary Table 9.

**Generation of the TKO cell line.** V6.5 cells were infected using lentiviruses harbouring the c3GIC9 vector (TRE3G-Cas9-P2A-GFP-PGK-Puro-IRES-rtta; a gift from L. Dow) containing gRNA/s targeting either *KLF4* only or *KLF2*, *KLF4* and *KLF5* in tandem. Following infection, the cells were selected using puromycin (LifeTech, K210015) and the clonal populations were manually picked. CRISPR-Cas9 expression from these stable cell lines was induced by the addition of dox for 72 h (1:1,000 dilution of a 2 mg ml<sup>-1</sup> stock) and the KO efficiency of each clonal population was verified by western blotting. Successful KO clones were then used for subsequent experiments (qPCR, ChIP-seq, 3C and HiChIP) after induction with dox for the indicated time periods.

**Antibodies.** Complete information regarding providers, catalogue numbers and dilutions of the antibodies used can be found in Supplementary Table 10.

**3C-qPCR.** Cells ( $1 \times 10^6$ ) were lysed in 300  $\mu\text{l}$  lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl and 0.2% Igepal CA630 with protease inhibitors) on ice for 20 min

followed by centrifugation at 2,500g for 5 min at 4 °C. The pellets were washed in lysis buffer and resuspended in 50  $\mu\text{l}$  0.5% SDS for 10 min at 65 °C. Water (145  $\mu\text{l}$ ) and 10% Triton X-100 (25  $\mu\text{l}$ ) were added and incubated 15 min at 37 °C. MboI enzyme (100 U) and 25  $\mu\text{l}$  NEB buffer 2 were then added and incubated overnight at 37 °C. Following enzyme inactivation, the DNA was ligated overnight at 16 °C by adding 120  $\mu\text{l}$  NEB T4 ligase buffer, 100  $\mu\text{l}$  Triton X-100 (10%), 6  $\mu\text{l}$  BSA (20 mg l<sup>-1</sup>), 100  $\mu\text{l}$  ATP (10 mM) and 5  $\mu\text{l}$  T4 ligase (NEB, M0202). The samples were treated with proteinase K and reverse crosslinked overnight at 65 °C. After RNase treatment, phenol/chloroform extraction and DNA precipitation, the pellets were dissolved in 100  $\mu\text{l}$  of 10 mM Tris pH 8. The samples (100 ng of material) were then used for qPCR amplification. All primer sequences can be found in Supplementary Table 9.

**ChIP-seq.** ChIP-seq was performed as previously described<sup>65</sup>. Specifically, cells were crosslinked in 1% formaldehyde at room temperature for 10 min and quenched with 125 mM glycine for 5 min at room temperature. The cells were used for KLF4 ChIP ( $50 \times 10^6$  cells) and H3K27ac ChIP ( $10 \times 10^6$  cells). The cell pellets were washed twice in PBS and resuspended in 400  $\mu\text{l}$  lysis buffer (10 mM Tris pH 8, 1 mM EDTA and 0.5% SDS) per  $20 \times 10^6$  cells. The cells were sonicated in a Bioruptor device (30 cycles of 30 s on/off; high setting) and spun down at the maximum speed for 10 min at 4 °C. The supernatants were diluted five times with dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris pH 8 and 167 mM NaCl) and incubated overnight with the respective antibody with rotation at 4 °C. Protein G Dynabeads (Thermo Scientific) pre-blocked with BSA protein (100 ng per 10  $\mu\text{l}$  Dynabeads) were added (10  $\mu\text{l}$  blocked Dynabeads per  $10 \times 10^6$  cells) the following day and incubated for 2–3 h at 4 °C. The beads were immobilized on a magnet and washed twice in low-salt buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 150 mM NaCl and 20 mM Tris pH 8), twice in high-salt buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 500 mM NaCl and 20 mM Tris pH 8), twice in LiCl buffer (0.25 M LiCl, 1% NP-40, 1% deoxycholic acid (sodium salt), 1 mM EDTA and 10 mM Tris pH 8) and once in TE buffer. The DNA was then eluted from the beads by incubating with 150  $\mu\text{l}$  elution buffer (1% SDS and 100 mM NaHCO<sub>3</sub>) for 20 min at 65 °C (vortexing every 10 min). The supernatants were collected and reverse crosslinked by incubation overnight at 65 °C in the presence of proteinase K. After RNase A treatment for 1 h at 37 °C, the DNA was purified using a minElute kit (Qiagen). The immunoprecipitated material (6–10 ng) was used for ChIP-seq library preparation using the KAPA Hyper prep kit (KAPA Biosystems). The libraries were sequenced on an Illumina HiSeq 2500 platform on SE50 mode.

**ATAC-seq.** Libraries were synthesized from 50,000 input cells as previously described<sup>66</sup>. The ATAC-seq libraries were first subjected to five cycles of pre-amplification. The library was assessed by quantitative PCR as previously described<sup>66</sup> to determine the suitable number of cycles required for the second round of PCR and the library was then PCR amplified for the appropriate number of cycles using Nextera primers. The samples were subjected to a dual size selection (0.55–1.5 $\times$ ) using SPRI beads (Beckman Coulter, B23317). Finally, the ATAC libraries were sequenced on a HiSeq 2500 platform in PE50 mode.

**RNA-seq.** Total RNA was prepared with TRIzol (Life Technologies, cat. no. 15996018) following the manufacturer's instructions. Libraries were generated by the Weill Cornell Genomics core facility using an Illumina TruSeq stranded mRNA library preparation kit (cat. no. 20020594) and sequenced on an Illumina HiSeq 4000 platform on SE50 mode.

**HiChIP.** HiChIPs were performed as previously described<sup>36</sup> with some modifications. Briefly, up to  $15 \times 10^6$  crosslinked cells (for the KLF4 HiChIPs, two samples of  $15 \times 10^6$  cells were combined at the end for each sample replicate) were resuspended in 500  $\mu\text{l}$  ice-cold HiC lysis buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 0.2% NP-40 and 1 $\times$ protease inhibitors) and rotated at 4 °C for 30 min. The nuclei were pelleted and washed once with 500  $\mu\text{l}$  ice-cold HiC lysis buffer. The pellet was then resuspended in 100  $\mu\text{l}$  of 0.5% SDS and incubated at 62 °C for 10 min. Water (285  $\mu\text{l}$ ) and 10% Triton X-100 (50  $\mu\text{l}$ ) were added, and the samples were rotated at 37 °C for 15 min. NEB buffer 2 (50  $\mu\text{l}$ ) and 25  $\mu\text{l}$  MboI restriction enzyme (15  $\mu\text{l}$ ; NEB, R0147) were then added and the sample was rotated at 37 °C for 2 h. The MboI was then heat inactivated at 62 °C for 20 min. We added 52  $\mu\text{l}$  incorporation master mix: 37.5  $\mu\text{l}$  of 0.4 mM biotin-dATP (Thermo Fisher, cat. no. 19524016); 4.5  $\mu\text{l}$  of a dCTP, dGTP, and dTTP mix at 10 mM each; and 10  $\mu\text{l}$  of 5 U  $\mu\text{l}^{-1}$  DNA polymerase I, large (Klenow) fragment (NEB, M0210). The reactions were then rotated at 37 °C for 1 h. Ligation master mix (150  $\mu\text{l}$  10 $\times$ NEB T4 DNA ligase buffer with 10 mM ATP (NEB, B0202), 125  $\mu\text{l}$  10% Triton X-100, 3  $\mu\text{l}$  50 mg ml<sup>-1</sup> BSA (Thermo Fisher, AM2616), 10  $\mu\text{l}$  400 U  $\mu\text{l}^{-1}$  T4 DNA ligase (NEB, M0202) and 660  $\mu\text{l}$  water; 948  $\mu\text{l}$ ) was then added. The reactions were next rotated at room temperature for 4 h. After proximity ligation, the nuclei were pelleted and the supernatant was removed. Nuclear lysis buffer (50 mM Tris-HCl pH 7.5, 10 mM EDTA, 0.5% SDS and 1 $\times$ protease inhibitors (Roche, cat. no. 11697498001)) was added to the nuclear pellet to a volume of 880  $\mu\text{l}$  and sonicated with a Bioruptor 300 (Diagenode) for eight cycles of 30 s each on a medium setting. The clarified samples were transferred to Eppendorf tubes and diluted

five times with ChIP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 7.5 and 167 mM NaCl). The cells were pre-cleared with 30  $\mu$ l Protein G Dynabeads (Life Technologies, 10004D) with rotation at 4 °C for 1 h. The supernatants were transferred into fresh tubes, antibody was added (8  $\mu$ g KLF4 antibody or 3  $\mu$ g H3K27Ac antibody for  $15 \times 10^6$  cells) and incubated overnight at 4 °C. The next day, 30  $\mu$ l Protein G Dynabeads were added to the samples and rotated at 4 °C for 2 h. After bead capture, the beads were washed three times each with low-salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 7.5 and 150 mM NaCl), high-salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 7.5 and 500 mM NaCl) and LiCl wash buffer (10 mM Tris-HCl pH 7.5, 250 mM LiCl, 1% NP-40, 1% sodium deoxycholate and 1 mM EDTA; freshly prepared). The samples were eluted with 150  $\mu$ l freshly prepared DNA elution buffer (50 mM sodium bicarbonate, pH 8.0 and 1% SDS) and incubated at 37 °C for 30 min with rotation. The supernatant was transferred to a fresh tube and elution was repeated with another 150  $\mu$ l elution buffer. Proteinase K (5  $\mu$ l at 20 mg ml<sup>-1</sup>; Thermo Fisher) was added to the 300  $\mu$ l reaction and the samples were incubated overnight at 65 °C. The samples were purified with DNA clean and concentrator columns (Zymo Research) and eluted in 10  $\mu$ l water. The post-ChIP DNA was quantified using a Qubit (Thermo Fisher) to estimate the amount of Tn5 (Illumina) needed to generate libraries with the correct size distribution (see below). Streptavidin C-1 beads (5  $\mu$ l; Thermo Fisher) were washed with Tween wash buffer (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl and 0.05% Tween-20) and then resuspended in 10  $\mu$ l 2 $\times$ biotin binding buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 2 M NaCl). Beads were added to the samples and incubated at room temperature for 15 min with shaking. After capture, the beads were washed twice by adding 500  $\mu$ l Tween wash buffer and incubating at 55 °C for 2 min with shaking. The samples were then washed in 100  $\mu$ l 1 $\times$ TD buffer (2 $\times$ TD buffer is 20 mM Tris-HCl pH 7.5, 10 mM magnesium chloride and 20% dimethylformamide). The beads were then resuspended in 25  $\mu$ l 2 $\times$ TD buffer, Tn5 (we used 2.5  $\mu$ l Tn5 for 50 ng post-ChIP DNA) and water to a final volume of 50  $\mu$ l. The amount of Tn5 was adjusted linearly for different amounts of post-ChIP DNA, with a maximum amount of 4  $\mu$ l Tn5. The samples were next incubated at 55 °C with interval shaking for 10 min. After removing the supernatant, 50 mM EDTA was added to the samples and incubated with interval shaking at 50 °C for 30 min. The beads were then washed two times each in 50 mM EDTA and Tween wash buffer at 55 °C for 2 min. Finally, the beads were washed in 10 mM Tris before PCR amplification. The beads were resuspended in 25  $\mu$ l 2 $\times$ Phusion HF (New England Biosciences), 1  $\mu$ l each Nextera Ad1\_noMX and Nextera Ad2.X at 12.5  $\mu$ M, and 23  $\mu$ l water. The following PCR program was performed: 72 °C for 5 min, 98 °C for 1 min, followed by cycling at 98 °C for 15 s, 63 °C for 30 s and 72 °C for 1 min (the cycle number was estimated based on the amount of material from the post-ChIP Qubit quantification—six cycles were used for approximately 50 ng of material, seven cycles for 25 ng, eight cycles for 12.5 ng, and so on). Size selection was performed using a two-sided size selection with Ampure XP beads. Following PCR, the libraries were placed on a magnet and eluted into new tubes. Ampure XP beads were added (25  $\mu$ l) and the supernatant was kept to capture fragments smaller than 700 bp. The supernatant was transferred to a new tube and 15  $\mu$ l of fresh beads was added to capture fragments larger than 300 bp. After size selection, the libraries were quantified using a Qubit and sent for analysis on a Bioanalyzer to check the quality and final size of the library. The libraries were sequenced on a HiSeq 2500 platform on PE75 mode.

**HiC.** The HiC experiments were performed starting with  $2 \times 10^6$  MEFs or iPSCs using the Arima-HiC kit as per the manufacturer's instructions.

**4C-seq.** Cells ( $10 \times 10^6$ ) were fixed following our ChIP-seq protocol (see above). The cells were lysed in 1 ml lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM EDTA, 1 $\times$ complete protease inhibitor, 0.5% NP-40 and 1% Triton X-100) on ice for 15 min. Following centrifugation at 2,500g for 5 min at 4 °C, the pellet was resuspended in 360  $\mu$ l Milli-Q water, 60  $\mu$ l 10 $\times$ DpnII restriction buffer and 15  $\mu$ l 10% SDS, and incubated at 37 °C for 1 h. Next, 150  $\mu$ l of 10% Triton X-100 was added and incubated for an additional 1 h. DpnII enzyme (4  $\mu$ l; NEB, R0543M) was then added and the samples were incubated overnight at 37 °C. Following enzyme inactivation, the samples were diluted with 4,860  $\mu$ l Milli-Q water, 700  $\mu$ l ligation buffer (500 mM Tris pH 7.5, 100 mM dithiothreitol, 100 mM MgCl<sub>2</sub> and 10 mM ATP) and 750  $\mu$ l Triton X-100, and incubated at 37 °C for 1 h. Ligase (2  $\mu$ l; NEB M0202M) was then added and the samples were ligated overnight at 16 °C. The samples were then treated with proteinase K and reverse crosslinked overnight. Following RNase treatment, phenol/chloroform extraction and DNA precipitation, the pellets were dissolved in 150  $\mu$ l of 10 mM Tris pH 7.5 and digested overnight at 37 °C by adding 50  $\mu$ l 10 $\times$ buffer B (Fermentas), 5  $\mu$ l Csp61 (Fermentas, ER0211) and 299  $\mu$ l Milli-Q water. Following enzyme inactivation, the samples were diluted in 12 ml Milli-Q water, 3  $\mu$ l ligase (NEB, M0202M) and 1.4 ml 10 $\times$ ligation buffer (500 mM Tris pH 7.5, 100 mM dithiothreitol, 100 mM MgCl<sub>2</sub> and 10 mM ATP) and incubated overnight at 65 °C. The DNA was purified by phenol/chloroform extraction, ethanol precipitation and Zymo columns (D4014). Finally, 150 ng DNA per reaction was used to PCR amplify the libraries using the KAPA HiFi enzyme (KAPA Biosystems, cat. no. 07958927001). All of the primer sequences can be

found in Supplementary Table 9. The samples were sequenced on a HiSeq 4000 platform in SE50 mode.

**RIME.** RIME was performed on three replicates for KLF4 and two for IgG, as previously described<sup>48</sup> with minor modifications. V6.5 cells ( $50 \times 10^6$ ) cultured in 2i conditions were used for each replicate. The cells were fixed, lysed, sonicated and incubated with the respective antibody-bound beads using the same conditions that were used for KLF4 ChIP-seq (see above). The samples were then washed ten times in RIPA buffer (50 mM HEPES pH 7.6, 1 mM EDTA, 0.7% (wt/vol) sodium deoxycholate, 1% (vol/vol) NP-40 and 0.5 M LiCl) and five times in 100 mM AMBIC solution. On-bead digests with trypsin (Promega) were carried out as in the original protocol. Peptides were acidified with 100% formic acid to a final concentration of 2% and then desalted using C18 resin self-packed STAGE Tip micro-columns.

**Co-immunoprecipitation and western blotting.** V6.5 cells ( $50 \times 10^6$ ) cultured in 2i conditions were collected for each co-immunoprecipitation experiment and resuspended in 0.5 ml lysis buffer (50 mM Tris pH 7.5, 100 mM NaCl, 0.2% Triton X-100, 0.5% glycerol and protease inhibitors). The cells were incubated on ice for 40 min followed by three cycles of sonication in a Bioruptor device (30 s on/off; high setting) and spun down at maximum speed for 10 min at 4 °C. The supernatants were diluted with additional lysis buffer to a final volume of 2 ml. The lysates were pre-cleared with 10  $\mu$ l Protein G Dynabeads (Thermo Scientific) for 30 min with rotation at 4 °C. The supernatant was then incubated with KLF4 or IgG antibodies for 2.5 h with rotation at 4 °C. Protein G Dynabeads (30  $\mu$ l) that had been pre-blocked with BSA were added to the samples and incubated for 1.5 h with rotation at 4 °C. Two washes were performed with lysis buffer followed by three washes with high-salt buffer (same as the lysis buffer but with 250 mM NaCl). Finally, the samples were eluted in loading buffer by boiling for 5 min and transferring the supernatant to a new tube. Western blots were then performed using the antibodies listed in Supplementary Table 10.

**ATAC-seq data mapping, peak calling and peak processing.** Paired-end reads were aligned to mm10 (bowtie2 version 2.3.2 (ref. 67)); -no-unal-local-very-sensitive-local-no-discordant-no-mixed-contain-overlap-dovetail -I 10 -X 2000) and the mitochondrial DNA alignments were excluded. The fragments marked as positional duplicates (sambamba version 0.6.6) or overlapping with mouseENCODE blacklisted genomic regions<sup>68</sup> (liftOver to mm10) were filtered out. The read ends were adjusted for Tn5 transposase offsets. Peaks were called at  $P < 10^{-5}$  (MACS version 2.1.1) per replicate and only common peaks between two independent biological replicates were retained for further analysis.

**ChIP-seq data analysis. Mapping, peak calling and peak processing.** The study and published ChIP-seq reads were trimmed for adaptors (cutadapt version 1.8.1) and low-quality ends (sickle version 1.33), respectively. Alignment to the mouse reference genome version mm10 (GRCm38.p4) was performed using bowtie2 version 2.3.2 (one mismatch in the seed alignment)<sup>67</sup>. Reads marked as positional duplicates (sambamba version 0.6.6) or overlapping with mouseENCODE blacklisted genomic regions (liftOver to mm10) were filtered out. The study ChIP-seq peaks (enrichment of signals over background determined by input samples) were called at  $P < 0.01$  (MACS version 2.1.1) per biological replicate and peaks detected in more than half of the biological replicates were retained for further analysis. Published ChIP-seq replicates were merged and peaks were called at  $P < 10^{-5}$  using input samples where applicable.

**Overlap analysis of ChIP-seq peaks for chromatin states of reprogramming cell types.** The chromatin states (1 kb resolution) during reprogramming were retrieved from ref. 8 and cis-regulatory elements were annotated from the chromatin states as in the original publication. The assignment of ChIP-seq peaks to cis-regulatory elements was determined by the largest degree of overlap in base pairs.

**RNA-seq and ChIP-seq gene ontology enrichment analysis.** The spatial proximity of ChIP-seq peaks to TSSs and enriched gene ontologies were uncovered utilizing GREAT (version 3.0.0). We selected the 'basal plus extension rule' for the association of gene-ontology annotations with regulatory domains (customized setting: 5 kb upstream and 1 kb downstream of the TSSs, and extended both directions by a further 250 kb). The enrichment of the ontology annotations was assessed by the binomial test of ChIP-seq peak overlaps with annotated regulatory regions. The DAVID knowledgebase<sup>69</sup> was used for pathway and biological process enrichment analysis for differentially expressed genes and gene groups (Supplementary Fig. 4e).

**Motif analysis.** We generated a random background (by shuffling the peaks randomly throughout the genome) for each KLF4 cluster to test the motif enrichment within each cluster. Analysis of the KLF4 clusters was performed with the use of HOMER and the 'findMotifsGenome.pl' command with the following parameters: '-bg random.bed -size 200 -len 15'. Only motifs with  $P \leq 1 \times 10^{-5}$  were considered significant. Two heat maps with the  $z$ -transformed '-log<sub>10</sub>[P value]' and

z-transformed 'motif frequency' of select motifs for each cluster are presented in Supplementary Fig. 1e.

#### Principle component analysis of the ATAC-, RNA- and ChIP-seq experiments.

We merged all accessible regions per H3K27ac peak detected in any reprogramming stage using bedtools v2.25.0. We calculated the coverage of the reads for each merged accessible region per H3K27ac peak for each replicate independently. For the RNA-seq data, we calculated the coverage for each exon and only exons with at least one read covering every single base of the exon were used for downstream analysis. Principle component analysis was performed with R.

**RNA-seq data analysis.** The expression of genes was quantified in TPM using quasi mapping (Salmon, version 0.8.2) to GENCODE (version M6, mm10). Salmon provides alignment-free transcript quantification information in a single step<sup>70</sup>.

**Processing of HiChIP and HiC datasets.** The HiChIP and HiC datasets were uniformly pre-processed using the HiC-bench platform<sup>71</sup>. Read pairs with a low mapping quality of MAPQ < 20, read pairs resulting from self-ligated fragments and short-range interactions resulting from read pairs aligning within 10 kb were filtered out prior to downstream analysis (together called ds.filtered). Counts-matrices per chromosome were created at a 10 kb bin size. For the HiChIP samples, trajectories of each matrix bin to both anchors were overlaid with the ChIP-seq signal of the respective sample, requiring a minimal overlap of 1 bp between a HiChIP-bin and a ChIP-peak. Only loops of which more than one anchor was supported by a ChIP-peak were kept for further analyses. Next, we applied sequencing-depth normalization (CPM) per replicate followed by statistical evaluation of the significant loops. We adapted the approach described in Mango<sup>72</sup> by performing a binomial test in each diagonal of the counts matrix up to a maximum distance of 2 MB.

High-confidence HiChIP loops were identified with a  $P < 0.1$  and CPM > 3 per loop across all replicates of a single condition. For high-confidence HiC loops, we adjusted those thresholds to avoid excessive noise, and applied filters of  $P < 0.01$  and CPM > 15 across all replicates of a single condition.

**Principal component analysis of HiChIP samples.** Principle component analysis was performed on each high-confidence loop detected from any sample; the per replicate normalized CPM was extracted before filtering for significant loops. Principle component analysis was performed using the prcomp function of R.

**Differential loop analysis.** Differential looping analysis was performed on each significant loop independently by applying an unpaired two-sided Student's *t*-test on the normalized counts (CPM) calculated before identifying significant loops between any pairwise comparisons: PSC-KLF4 versus day 3-KLF4, PSC-KLF4 versus day 6-KLF4, day 3-KLF4 versus day 6-KLF4, PSC-H3K27ac versus MEF-H3K27ac, TKO-0h versus TKO-24h. We calculated the  $\log_2[\text{FC}]$  between the average CPMs per condition for the same pairwise comparisons after adding a pseudo-count of one to each replicate and loop. For constant H3K27ac loops in MEF versus PSC or TKO-0h versus TKO-24h, we selected loops using a  $P > 0.5$  and absolute  $\log[\text{FC}] < 0.5$ . MEF- and PSC-specific H3K27ac loops were selected using a  $P < 0.1$  and  $\log[\text{FC}] > 2$  and  $\log[\text{FC}] < -2$ , respectively. The TKO-0h- and TKO-24h-specific loops were selected using a  $P < 0.05$  and  $\log[\text{FC}] > 0.58$  and  $\log[\text{FC}] < -0.58$ , respectively. Mid- and late-established KLF4 loops were selected by applying a  $P < 0.01$  and  $\log[\text{FC}] > 2$  in the pairwise comparisons of PSC-KLF4 versus day 3-KLF4, day 6-KLF4 versus day 3-KLF4 (mid), PSC-KLF4 versus day 3-KLF4 and PSC-KLF4 versus day 6-KLF4 (late). Early-lost and mid-lost KLF4 loops were selected by applying a  $P < 0.01$  and  $\log[\text{FC}] < -2$  in the pairwise comparisons of PSC-KLF4 versus day 6-KLF4, PSC-KLF4 versus day 3-KLF4 (early-lost), PSC-KLF4 versus day 3-KLF4 and day 6-KLF4 versus day 3-KLF4 (mid-lost). For the differential comparison of significant HiC loops, we applied distance-normalization as previously described<sup>73</sup> before calculating significance and fold-changes between the PSC and MEF HiC loops. Differential HiC loops were selected by applying a  $P < 0.1$  and  $\log[\text{FC}] < -0.32$  or  $\log[\text{FC}] > 0.32$  in the pairwise comparison of PSC HiC versus MEF HiC. All calculations were performed in R version 3.3.0.

The filtering of differential HiChIP loops was done without correcting for multiple testing (see the 'Statistics and reproducibility' section for details).

#### Annotation of H3K27ac HiChIP loop anchors as promoters or enhancers.

H3K27ac HiChIP loop anchors were overlapped with the TSSs of GENCODE (version M6) protein-coding genes. The presence of one or more TSS was considered a promoter HiChIP anchor and the absence of any TSS but presence of at least one H3K27ac ChIP-seq constituted an enhancer HiChIP anchor. All HiChIP anchors, either promoter, enhancer or otherwise desolate, were considered in estimating connectivity.

**RNA expression integration with differential HiChIP loops.** We overlapped canonical TSSs of protein-coding genes (transcript support level (TSL) = 1) downloaded from Ensembl Genes V85 for the mouse genome mm10 with all loop

anchors. We filtered genes by occurrence of differential loop clusters obtained from the HiChIP experiments with TPM > 1 in at least one reprogramming stage. For H3K27ac HiChIP data integration, we assigned genes to MEF/PSC-specific/constant loops if their TSSs were found in  $\geq 1$  MEF/PSC-specific/constant loops but in none of the other categories. We applied an unpaired, one-sided Student's *t*-test between the  $\log[\text{FC}]$  of genes of constant H3K27ac loops and genes with MEF/PSC-specific loops, following the alternative hypothesis of a positive correlation between looping and expression changes (null hypothesis: MEF/PSC-specific H3K27ac contacts have no or negative correlation with the gene expression of the associated genes). We also compared the  $\log[\text{FC}]$  of genes of constant loops with all annotated protein-coding genes using a two-sided unpaired Student's *t*-test. We followed the same approach for the integration of expression data with differential loops from Klf-TKO H3K27ac HiChIP experiments.

#### Coregulation of gene expression by H3K27ac HiChIP enhancer hubs.

Promoter anchors of enhancer-mediated loops were filtered for protein-coding genes with TPM > 1 in PSC. Enhancer fragments that were in contact with between two and ten promoter fragments in PSC-specific H3K27ac loops were selected. Genes were paired across different promoter fragments connecting to theWe applied an unpaired, one-sided same enhancer anchor (called 'hub') and repeated gene pairs were removed from the overall pool. Gene pairs were considered co-expressed if both genes were upregulated in PSCs compared with MEFs (adjusted  $P < 10^{-2}$  and  $\log[\text{FC}] > 1$ ) and vice versa for downregulation. Alternately, a pair with at least one gene unchanged between MEFs and PSCs constituted an unchanged gene pairs. Statistical enrichment of coregulated gene pairs in the hubs was performed using a Fisher's exact test. The background probability was calculating using an equal number of random gene pairs (protein-coding genes that have an expression TPM > 1 in PSCs) either of similar linear distance with our test group (global random) or within the same TADs. Recently published software<sup>73</sup> was used to call TADs from normalized corrected HiC matrices in PSCs processed at a resolution of 10 kb with the use of the following parameters: -minDepth 120000-maxDepth 420000-thresholdComparison 0.001-delta 0.01-correctForMultipleTesting fdr.

#### ChIP-seq enrichment at low or highly connected H3K27ac PSC-specific enhancer anchors.

H3K27ac HiChIP enhancer anchors were selected for low ( $n = 1,183$ ) or high connectivity (four or more anchors;  $n = 1,014$ ). LOLA analysis<sup>74</sup> was performed in these two groups of ChIP-seq peaks.

**KLF4 looping involved in RNA expression changes.** We selected expressed genes within anchors of each KLF4 loop cluster, and further filtered for differentially expressed genes between PSCs and day 3 (DESeq results: FDR < 0.01;  $\log[\text{FC}] > 1.0$  (upregulated) or  $\log[\text{FC}] < -1.0$  (downregulated)). Genes determined as 'no change' were selected by applying an FDR > 0.5 and absolute  $\log[\text{FC}] < 0.25$ .

**LOLA enrichment analysis.** Anchors of each differential loop were mapped back to the original ChIP or ATAC-peaks. Each anchor that overlapped an actual ChIP or ATAC-peak by at least 1 bp was subjected to further analysis. Next, we applied LOLA version 1.8.0 (ref. <sup>74</sup>) against the LOLA region databases (regionDB) for mm10 and additional published datasets (see Supplementary Table 12). We excluded all ChIP-seq datasets that were marked as treated with any agent and had fewer than 3,000 peaks in total, and selected the dataset with the highest number of peaks when multiples existed for the same factor and cell line. As a universe for LOLA, we used only unique ChIP or ATAC-peaks from the union of all ChIP or ATAC-seq peaks for the respective antibody across all reprogramming stages.

**Virtual 4C.** We used filtered HiChIP read pairs as described above before binning and normalizing each replicate. We extracted read pairs for which a read mate maps within  $\pm 10$  kb around the virtual viewpoint. We defined successive overlapping windows for each chromosome at a 10 kb resolution, overlapping by 95% of their length. We then counted the second mapped read mate in all overlapping bins. Read counts were normalized to the total sequencing depth of the respective replicate using edgeR<sup>75</sup> version 3.14.0 to calculate the CPM per bin. Significant differences between TKO-0h and TKO-24h or MEFs and PSCs were calculated using the edgeR function glmQLFTest (we did not correct for multiple testing because the overlapping windows result in dependent data and thus a correction might be too conservative in reducing the statistical power). Visualization was done using average CPM signals per condition.

**Analysis of 4C-seq data.** The 4C-seq data were analysed in a similar fashion as recently described<sup>76</sup>. Viewpoint primers were trimmed off from all sequencing reads using seqtk (version 1.3.0). The read-sequence was aligned using bowtie v1.0.0 against a reduced genome that consists only of reference genome sequences adjacent to DpnII cut-sites (following the 4C-ker pipeline<sup>76</sup>). Next, the genome was binned into 10 kb bins shifted by 500 bp (95% overlap with the adjacent bins). Reads were counted by unique alignment position in all overlapping bins. The read counts per bin were normalized to the sequencing depth per replicate using edgeR<sup>75</sup> (version 3.14.0), resulting in CPM. Visualization was done using average CPM signals per condition.

**Statistics and reproducibility.** All sequencing experiments and functional analyses involved at least two biological replicates, that is, either independently prepared, cultured, treated and isolated cell lines (ESCs, iPSCs, MEFs and reprogramming intermediates) or different ESC clones after genetic manipulation. Statistical analyses for differential gene expression and differential ChIP-seq peaks was conducted with DESeq2 using two biologically independent replicates. Gene-ontology and pathway-enrichment significance was calculated with a modified Fisher's exact test, whereas motif and transcription-factor enrichment significance was evaluated with the use of hypergeometric test. Differential HiChIP contacts were called based on an unpaired two-sided Student's *t*-test without correction for multiple testing. However, 83.2–100% of our called differential loops were also detected following an independent approach using the R Bioconductor package edgeR<sup>75</sup> (after intra-sample sequencing depth normalization with the 'cpm' function, inter-sample dispersion correction using the 'estimateCommonDisp' and 'estimateTagwiseDisp' functions, and 'glmQLFit' and 'glmQLFTest' for differential analysis followed by Bonferroni–Hochberg correction for multiple testing reporting the FDR) and applying the same thresholds (FDR < 0.1 and FC > 2 for PSC/MEF-specific loops, FDR < 0.05 and FC > 1.5 for Klf-TKO HiChIP data, and FDR < 0.01 and FC > 2 for Klf4-associated HiChIP data; See Supplementary Table 11). The independent approach confirms stringent statistical filtering and a low rate of false-positive differential loops. Statistical significance of the differences in the expression or connectivity levels among different groups of genes and anchors (Fig. 3a,b,d and Supplementary Figs. 3b, 4b, 5d, 6d) was calculated using two-tailed Wilcoxon rank-sum test.

When we expected changes towards one direction we used a one-tailed Student's *t*-test under the following null hypotheses for Figs. 2b, 6c and Supplementary Fig. 2d:

$H_0$ : The mean change in expression of genes associated with differential loops (for example, expression fold change of PSC/MEFs in PSC-specific loops) is unchanged or follows a negative correlation with looping strength changes (for example,  $\mu_{\text{PSC-specific loops}} \geq \mu_{\text{constant loops}}$ ).

$H_A$ : The mean change in expression of genes associated with differential loops (for example, expression fold change of PSC/MEF in PSC-specific loops) follows a positive correlation with looping strength changes (for example,  $\mu_{\text{PSC-specific loops}} \geq \mu_{\text{constant loops}}$ ).

Similarly, for comparisons of the expression levels determined by RT-qPCR and 3C-qPCR between mutant or KO clones and WT cells (Figs. 4f, 5f,g, 7d,e,i,j and Supplementary Figs. 4i,j, 6b and 7d), we used a one-tailed Student's *t*-test under the following hypothesis:

$H_0$ : The mean expression/interaction score of genes/enhancers in mutant (or KO) cells is greater than or equal to the expression or interaction score of the same genes/enhancers in WT cells.

$$\mu_{\text{KO/mut}} \geq \mu_{\text{WT}}$$

$H_A$ : The mean expression or interaction score of genes or enhancers in mutant (or KO) cells is smaller than the expression or interaction score of the same genes or enhancers in WT cells.

$$\mu_{\text{KO/mut}} < \mu_{\text{WT}}$$

A paired Student's *t*-test was performed when observations between the two conditions were paired and each pair of WT and KO or mutant clone was tested in independent experiments (Fig. 7d and Supplementary Figs. 4i,j and 6b). The

number of samples tested and the statistical test used in each case is described in the respective figure legends.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

All genomics data (RNA-seq, ChIP-seq, ATAC-seq and HiChIP) have been deposited in the Gene Expression Omnibus (GEO) under the accession code GSE113431. RIME data have deposited to the ProteomeXchange Consortium via the PRIDE partner repository under the identifier PXD014631 ([www.ebi.ac.uk/pride](http://www.ebi.ac.uk/pride)). The accession codes for all previously published datasets that were used in this study at listed in Supplementary Table 12. The source data for Figs. 4f, 6f,g, 7d,e,i,j and Supplementary Figs. 4i,j, 6b, 7d have been provided as Supplementary Table 11.

### Code availability

The computational code for the processing of HiChIP/HiC is available under <https://github.com/NYU-BFX/hic-bench>. All other custom computational code are available from the corresponding author on request.

### References

- Liu, Y et al. Widespread mitotic bookmarking by histone marks and transcription factors in pluripotent stem cells. *Cell Rep.* **19**, 1283–1293 (2017).
- Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1–21.29.9 (2015).
- Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
- The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
- Lazaris, C., Kelly, S., Ntziachristos, P., Aifantis, I. & Tsirigos, A. HiC-bench: comprehensive and reproducible Hi-C data analysis designed for parameter exploration and benchmarking. *BMC Genom.* **18**, 22 (2017).
- Gong, Y. et al. Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. *Nat. Commun.* **9**, 542 (2018).
- Ramirez, F. et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* **9**, 189 (2018).
- Sheffield, N. C. & Bock, C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and bioconductor. *Bioinformatics* **32**, 587–589 (2016).
- Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
- Raviram, R. et al. 4C-ker: a method to reproducibly identify genome-wide interactions captured by 4C-seq experiments. *PLoS Comput. Biol.* **12**, e1004780 (2016).

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a | Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated
- Clearly defined error bars  
*State explicitly what error bars represent (e.g. SD, SE, CI)*

*Our web collection on [statistics for biologists](#) may be useful.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Published Hi-C and ChIPseq were downloaded using stratoolkit version 2.8.0. FACs data were collected using BD FACSDiva v8.0.3 Software

Data analysis

HiC-bench (<https://github.com/NYU-BFX/hic-bench>), bowtie2 version 2.2.3, GenomicTool, R version 3.3.0, LOLA version 1.8.0, sambamba version 0.6.6, MACS version 2.1.1, cutadapt version 1.8.1, GREAT version 3.0.0, GENCODE version M6, bedtools v2.25.0, Salmon version 0.8.2, cutadapt version 1.8.1, sickle version 1.33, tophat 2.1.1, HTSeq 0.5.4p3, homer 4.10, Juicebox, UCSC liftOver, R Bioconductor package edgeR version 3.14.0, bowtie version 1.0.0, seqtk version 1.3.0, DESeq2, FlowJo 9.3.2

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data (RNA-seq, ChIP-seq, ATAC-seq, HiC and HiChIP) were submitted in GEO under the accession code GSE113431. RIME data have deposited to the ProteomeXchange Consortium via the PRIDE partner repository under the identifier PXD014631. All data will be released upon formal acceptance of the paper.

## Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Data exclusions

Replication

Randomization

Blinding

## Reporting for specific materials, systems and methods

### Materials & experimental systems

n/a	Involvement in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Unique biological materials
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants

### Methods

n/a	Involvement in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Unique biological materials

Policy information about [availability of materials](#)

Obtaining unique materials

## Antibodies

Antibodies used

Antibody name/ Company /Catalog number /Dilution-concentration (see also Suppl. Table 11)  
 Thy1 ebioscience 48090282 for FACS 1ul in 200ul  
 SSEA1 biogend 125608 for FACS 1ul in 200ul  
 SSEA1 Miltenyl Biotech 130-094-530 for MACs manufacturer instructions  
 H3K27ac Abcam ab4729 for ChIP and HiChIP 3ug/10M cells

KLF4 R&D AF3158 for ChIP and HiChIP 3ug/10M cells, for WB: 1ul in 2.5ml  
 KLF5 R&D AF3758 for WB 1ul in 1ml  
 KLF2 Novus biologicals NBP6181 for WB 1ul in 500ul  
 ESRRB PPMX PPH6705 for WB 1ul in 1ml  
 NANOG Bethyl laboratories A300-397A for WB 1ul in 10ml  
 ACTIN Abcam ab49900 for WB 1ul in 5ml  
 IgG Calbiochem NI02 for IP 3ug/10M cells  
 BRD4 Bethyl laboratories A301-985A50 for WB 1ul in 1ml  
 MED1 Bethyl laboratories A300-793A for WB 1ul in 1ml  
 SMC1a Bethyl laboratories A300-055A for WB 1ul in 1ml  
 RING1b Bethyl laboratories A302-869A for WB 1ul in 1ml  
 SUZ12 Santa cruz sc46264 for WB 1ul in 1ml  
 LSD1 Abcam ab 17721 for WB 1ul in 1ml

## Validation

ChIP antibodies were validated by ChIP-qPCR and Western blots.  
 The specificity of KLF4 antibody and potential crossreactivity with related KLF2 and KLF5 factors was also validated by performing WB analysis in ESCs before or after depletion of KLF4 by CRISPR/Cas9.  
 FACs antibodies were validated by using negative and positive control cell lines (eg. MEF or ESCs) and unstained controls.  
 IP antibodies were tested by comparing with IgG controls.  
 WB antibodies were validated by comparing WT cells to Knockout cells (when available).  
 We did not check species specificity, since it was not relevant for our study.

## Eukaryotic cell lines

Policy information about [cell lines](#)

## Cell line source(s)

MEFs were isolated from embryos E13.5 from mice harboring an OKSM polycistronic cassette in the Col1a1 locus and an M2-rtTA in the Rosa26 locus (see Stadtfeld et al, Nature Methods 2010), ESC V6.5 were obtained from the Hochedlinger lab, 293T obtained from the Hochedlinger lab.

## Authentication

For MEFs genotyping PCR was used to validate the presence of all alleles. For ESC V6.5 previous experiments of tetraploid complementation and generation of offsprings with Agouti coat color was used to validate pluripotency. Morphology of ESC colonies, RT-qPCR and IF for pluripotency markers were routinely used for validation of their stemness. 293T have not been authenticated for this study.

## Mycoplasma contamination

All cell lines were routinely tested for mycoplasma contamination and they were negative.

Commonly misidentified lines  
(See [ICLAC](#) register)

Commonly misidentified cell lines were not used in this study.

## Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

## Laboratory animals

Crosses between the B6;129S4-Col1a1tm1(tetO-Pou5f1,-Klf4,-Sox2,-Myc)Hoch/ (JAX 011001) and B6.Cg-Gt(ROSA)26Sortm1(rtTA\*M2)Jae/J animals were used for generation of MEFs from male and female embryos at gestational age between E13.5 and E15.5.

## Wild animals

The study did not involve wild animals

## Field-collected samples

The study did not involve field-collected samples

## ChIP-seq

## Data deposition

- Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).  
 Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

## Data access links

*May remain private before publication.*

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE113431>  
 token : oholsiaytvujtpw

## Files in database submission

GSM3103910 ES-Klf4 rep1  
 GSM3103911 ES-Klf4 rep2  
 GSM3103912 ES-Klf4 rep3  
 GSM3103913 ES-Klf4 rep4  
 GSM3103914 d3-Klf4 rep1  
 GSM3103915 d3-Klf4 rep2  
 GSM3103916 d3-Klf4 rep3  
 GSM3103917 d6-Klf4 rep1

GSM3103918 d6-Klf4 rep2  
GSM3103919 d6-Klf4 rep3  
GSM3103920 d6-Klf4 rep4  
GSM3103921 ES-H3K27ac rep1  
GSM3103922 ES-H3K27ac rep2  
GSM3103923 MEF-H3K27ac rep1  
GSM3103924 MEF-H3K27ac rep2  
GSM3106249 ATAC-seq\_MEF\_1  
GSM3106250 ATAC-seq\_MEF\_2  
GSM3106251 ATAC-seq\_DAY3\_1  
GSM3106252 ATAC-seq\_DAY3\_2  
GSM3106253 ATAC-seq\_DAY6\_1  
GSM3106254 ATAC-seq\_DAY6\_2  
GSM3106255 ATAC-seq\_DAY9\_1  
GSM3106256 ATAC-seq\_DAY9\_2  
GSM3106257 ATAC-seq\_ESC\_1  
GSM3106258 ATAC-seq\_ESC\_2  
GSM3106259 EC-DG-3458-H3K27AC\_MEF\_1  
GSM3106260 EC-DG-3458-H3K27AC\_MEF\_2  
GSM3106261 EC-EA-3702-H3K27AC\_DAY3\_1  
GSM3106262 EC-DG-3458-H3K27AC\_DAY3\_1  
GSM3106263 EC-DG-3458-H3K27AC\_DAY3\_2  
GSM3106264 EC-DG-3458-H3K27AC\_DAY6\_1  
GSM3106265 EC-DG-3458-H3K27AC\_DAY6\_2  
GSM3106266 EC-EA-3702-H3K27AC\_DAY6\_1  
GSM3106267 EC-EA-3702-H3K27AC\_DAY9\_1  
GSM3106268 EC-EA-3702-H3K27AC\_DAY9\_2  
GSM3106269 EC-DG-3458-H3K27AC\_ESC\_1  
GSM3106270 EC-DG-3458-H3K27AC\_ESC\_2  
GSM3106271 EC-EA-3702-KLF4\_DAY3\_1  
GSM3106272 EC-DG-4040-KLF4\_DAY3\_1  
GSM3106273 EC-EA-3702-KLF4\_DAY6\_1  
GSM3106274 EC-DG-4040-KLF4\_DAY6\_1  
GSM3106275 EC-EA-3702-KLF4\_DAY9\_1  
GSM3106276 EC-EA-3702-KLF4\_DAY9\_2  
GSM3106277 EC-EA-2697-KLF4\_ESC\_1  
GSM3106278 EC-EA-2643-KLF4\_ESC\_1  
GSM3106279 EC-EA-3702-KLF4\_ESC\_2I\_1  
GSM3106280 EC-DG-3012-KLF4\_ESC\_2I\_1  
GSM3106281 EC-DG-3458-INPUT\_MEF  
GSM3106282 EC-EA-3702-INPUT\_DAY3  
GSM3106283 EC-DG-3458-INPUT\_DAY3  
GSM3106284 EC-DG-4040-INPUT\_DAY3  
GSM3106285 EC-DG-3458-INPUT\_DAY6  
GSM3106286 EC-EA-3702-INPUT\_DAY6  
GSM3106287 EC-DG-4040-INPUT\_DAY6  
GSM3106288 EC-EA-3702-INPUT\_DAY9  
GSM3106289 EC-DG-3458-INPUT\_ESC  
GSM3106290 EC-EA-2697-INPUT\_ESC  
GSM3106291 EC-EA-2643-INPUT\_ESC  
GSM3106292 EC-EA-3702-INPUT\_ESC\_2I  
GSM3106293 EC-DG-3012-INPUT\_ESC\_2I  
GSM3106294 RNA-seq\_MEF\_1  
GSM3106295 RNA-seq\_MEF\_2  
GSM3106296 RNA-seq\_D3\_1  
GSM3106297 RNA-seq\_D3\_2  
GSM3106298 RNA-seq\_D6\_1  
GSM3106299 RNA-seq\_D6\_2  
GSM3106300 RNA-seq\_D9\_1  
GSM3106301 RNA-seq\_D9\_2  
GSM3106302 RNA-seq\_ESC\_1  
GSM3106303 RNA-seq\_ESC\_2  
GSM3714370 TKO 0h A: H3K27ac WT ChIP-seq  
GSM3714371 TKO 0h B: H3K27ac WT ChIP-seq  
GSM3714372 TKO 24h A: H3K27ac KLF4 KO ChIP-seq  
GSM3714373 TKO 24h B: H3K27ac KLF4 KO ChIP-seq  
GSM3714374 Input H3K27ac TKO 1  
GSM3714375 Input H3K27ac TKO 2  
GSM3714376 iPSC KLF4 A ChIP-seq  
GSM3714377 iPSC KLF4 B ChIP-seq  
GSM3714378 iPSC H3K27ac A ChIP-seq  
GSM3714379 iPSC H3K27ac B ChIP-seq  
GSM3714380 Input iPSC H3K27ac  
GSM3714381 Input iPSC KLF4  
GSM3714382 TKO 0h A: RNA-seq in WT ESCs

GSM3714383 TKO 0h B: RNA-seq in WT ESCs  
 GSM3714384 TKO 24h A: RNA-seq in ESCs after KLF4 KO  
 GSM3714385 TKO 24h B: RNA-seq in ESCs after KLF4 KO  
 GSM3714977 TKO-0h H3K27ac rep1  
 GSM3714978 TKO-0h H3K27ac rep2  
 GSM3714979 TKO-0h H3K27ac rep3  
 GSM3714980 TKO-24h H3K27ac rep1  
 GSM3714981 TKO-24h H3K27ac rep2  
 GSM3714982 TKO-24h H3K27ac rep3  
 GSM3714983 ES-HiC-Arima rep1  
 GSM3714984 ES-HiC-Arima rep2  
 GSM3714985 MEF-HiC-Arima rep1  
 GSM3714986 MEF-HiC-Arima rep2

Genome browser session  
 (e.g. [UCSC](#))

No longer applicable

## Methodology

Replicates

At least two replicates were used for every ChIP-seq dataset and only peaks consistent in both replicates (or at least 2 in case of more replicates) were used.

Sequencing depth

Minimum of 10M.

Antibodies

H3K27ac Abcam ab4729 for ChIP and HiChIP 3ug/10M cells  
 KLF4 R&D AF3158 for ChIP and HiChIP 3ug/10M cells, for WB: 1ul in 2.5ml

Peak calling parameters

For read mapping we used default parameters of bowtie2 with (-N 1). For peak calling we used default parameters of macs2 with macs2 callpeak -g mm -p 1e-2 -B --SPMR --call-summits or macs2 callpeak -g mm -p 1e-3 --shift 37 --extsize 73 --call-summits

Data quality

In-house data was trimmed for adapter contamination. The percentage of peaks that are at FDR 5% and above 5-fold enrichment range between 8 % and 56 %

Software

bowtie2, GenomicTool, R version 3.3.0, LOLA version 1.8.0, sambamba version 0.6.6, MACS version 2.1.1, cutapt version 1.8.1, GREAT version 3.0.0, bedtools v2.25.0, sickle version 1.33, tophat 2.1.1, HTSeq 0.5.4p3, homer 4.10

## Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Cells were trypsinized and stained with indicated antibodies according to manufacturer's instructions. Posseld2 was used for isolation of SSEA1 positive cells

Instrument

BD FACS-Cantoll, MACs Militenyi Biotec

Software

FlowJo 9.3.2, BD FACSDiva v8.0.3

Cell population abundance

The purity of MACs sorted samples was validated by FACs analysis and varied between 95-98%.

Gating strategy

Relevant cells were first gated on Forward scatter (FSC) and Side Scatter (SSC), then doublets were excluded in FSC-H/W and live cells were selected as DAPI negative. Thy1 positive and SSEA1 positive cells were gated as indicated in Supplementary Figure 1a based on "unstained" sample as control.

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.